STATE OF CALIFORNIA DEPARTMENT OF TRANSPORTATION

TECHNICAL REPORT DOCUMENTATION PAGE

TR0003 (REV. 10/98)

1. REPORT NUMBER	2. GOVERNMENT ASSOCIATION NUMBER	3. RECIPIENT'S CATALOG NUMBER
CA11-1214		

4. TITLE AND SUBTITLE	5. REPORT DATE
Strategies for Reducing Pedestrian and Bicyclist Injury at the Corridor Level	JULY 2011
	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S)	8. PERFORMING ORGANIZATION REPORT NO.
David Ragland, Offer Grembek, Phyllis Orrick	UCB-ITS-PRR-XXXX-XX
9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. WORK UNIT NUMBER
UC Berkeley Safe Transportation Research & Education Center 2614 Dwight Way, #7374 Berkeley, CA 94720-7374	11. CONTRACT OR GRANT NUMBER 65A0273
12. SPONSORING AGENCY AND ADDRESS	13. TYPE OF REPORT AND PERIOD COVERED
California Department of Transportation Division of Research and Innovation, MS-83 1227 O Street Sacramento CA 95814	Final report 14. SPONSORING AGENCY CODE

15. SUPPLEMENTAL NOTES

16. ABSTRACT

Methods for identifying sites with potential for preventing traffic fatalities and injuries have been developed for *vehicle-vehicle* collisions. This study was funded by the California Department of Transportation (Caltrans) to develop methods for identifying sites where there is potential for significant reductions in *pedestrian and bicyclist* injury. Data from 1998-2007 from a 16.5-mile section of San Pablo Avenue (SR 123) in the San Francisco East Bay was used as a study area. Several approaches for identifying sites with high potential for reducing pedestrian and bicyclist injury were evaluated and compared, a framework was developed for conducting benefit-cost analyses, and a prototype was developed for a training protocol for conducting analyses of pedestrian and bicyclist safety in a corridor or network. The basic principle followed is that sites with the most potential for reducing injury are those sites where the most injuries can be prevented per dollar spent. Everything else being equal, these sites are the ones with the highest expected number of injuries if nothing is done. Prior history is typically used to make this estimate, but this may not be sufficient, especially if the underlying rates are low. Several approaches to developing statistically stable estimates are explained and compared: (i) extend the number years for both the baseline and follow-up periods, (ii) expand the size of the target sites considered, and (iii) apply Bayesian methods to include a modeled estimate of risk in the calculation. Strengths and weaknesses of each of these approaches are discussed with illustrations from the study area.

17. KEY WORDS	18. DISTRIBUTION STATEMENT	
Pedestrian injury; Bicyclist injury; Expected injuries; Benefit-cost analyses; High collision locations	No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22161	
19. SECURITY CLASSIFICATION (of this report)	20. NUMBER OF PAGES	21. PRICE
Unclassified	73	

STRATEGIES FOR REDUCING PEDESTRIAN AND BICYCLIST INJURY AT THE CORRIDOR LEVEL:

FINAL TECHNICAL REPORT

SECOND DRAFT

PREPARED BY THE

UC BERKELEY SAFE TRANSPORTATION RESEARCH AND EDUCATION CENTER

FOR THE

CALIFORNIA DEPARTMENT OF TRANSPORTATION

JULY 8, 2011

DAVID R. RAGLAND, PHD, MPH

OFFER GREMBEK, PHD

PHYLLIS ORRICK

DISCLAIMER STATEMENT

ACKNOWLEDGEMENTS

The authors would like to thank the California Department of Transportation for their support of this project. We especially acknowledge the support and guidance of Richard Haggstrom, Pedestrian Safety Program Leader at Caltrans until his retirement in December 2010. We also acknowledge SafeTREC graduate student researchers Tierra Bills, Terri O'Connor and SafeTREC staff members John Bigham and Swati Pande for their work on constructing data files and conducing analyses for the project.

Executive Summary

While roadway safety data systems and methods of analysis have generally focused on motor vehicle and transit performance and efficiency, walking and bicycling are emerging as viable alternative travel modes. Parallel data systems and methods of analysis for these travel modes are essential to incorporating them into this multi-modal transportation system.

For this study we used data from San Pablo Avenue, a major arterial along the east shore of the San Francisco Bay. This study, funded by the California Department of Transportation (Caltrans), develops methods to support multi-modal transportation, with a focus on identifying sites with potential for significant reductions in pedestrian and bicyclist injury.

The San Pablo Avenue SMART Corridor is a system of freeways and major arterials serving the east shore of the San Francisco Bay, from downtown Oakland north to the City of Hercules. The corridor utilizes intelligent transportation system (ITS) technologies to increase and enhance transportation mobility throughout several East Bay communities.

Our previous research suggests that pedestrian and bicyclist events are clustered in specific locations and that there is considerable variation over time, which is more pronounced at individual sites because the number of events is small, making it difficult to accurately predict where a collision will occur. This, in turn results complicates directing resources where they are most likely to prevent collisions. Identifying high-risk collision locations has generally focused on individual locations and used past history to generate an expected number of collisions.

To effectively utilize limited resources, we consider four approaches to improve the efficiency of selecting sites: (i) increasing the time horizon for events, either years of history or post-treatment follow-up, (ii) increasing the geographic scale (from specific sites to corridors, zones, or an entire network), (iii) combining sites with similar characteristics and (iv) creating estimates using a Bayesian method which combined on prior knowledge with the realization of crashes.

To calculate benefit-cost (BC), a method is required that includes, at a minimum, the following specifications: (i) compares across different levels of scale; (ii) considers different levels of injury; and (iii) compares different types of injury (e.g., pedestrian, bicyclist, motor vehicle occupant).

Within the context of this study we developed a stand-alone tool based on an approach that uses Crash Reduction Factors (CRFs) which can be applied differentially to the various collisions occurring at a site, a set of sites, a corridor, or a zone. The tool uses standard formulas for benefit-cost calculation, which are included in the Highway Safety

Improvement Plan (HSIP) program guide and also links to extensive HSIP safety resources.

A database including all Vulnerable Road User (VRU) collisions was constructed using the Statewide Integrated Traffic Records System (SWITRS). The database includes all (413) pedestrian and bicyclist collisions from 1998 to 2007, geocoded and matched to the closest intersection.

The conclusions from this study are:

- Basing decisions on individual intersections and single years is of limited efficacy and will yield substantial numbers of two types of errors: (i) false positives, i.e., selecting a particular intersection for treatment when in fact there would be a fairly small number of collisions there in subsequent years and (ii) false negatives, i.e., failing to select an intersection when in fact that intersection would have a significant number of collisions in ensuing years. If, as is the case, the goal is to prevent the maximum number of collisions per dollar spent, prioritizing spending according to annual crashes at individual intersections would yield limited success.

- Confidence intervals around crash frequencies decrease most rapidly over the first three years, and then decrease more slowly. This may indicate that using a sampling period of three years, as is common, provides a good balance between reducing statistical variation and accounting for changes in the intersections over time. However, for low-crash intersections a longer period, say five years, may be necessary for achieving more stability in estimating the expected number of crashes.

- There are high concentrations of collisions not only for individual intersections, but high concentrations within spatial clusters of intersections, suggesting there are factors affecting not just the individual intersection, but affecting entire sets of intersections. There is an advantage in terms of increased stability of estimated future collisions of clustering adjacent intersections over focusing on individual intersections.

- The table below summarizes how well two the baseline and outcome periods conform for the different approaches evaluated using the *CV(RMSE)* measure. The results demonstrate that increasing the compared intervals to 5 years and grouping together 5 adjusted intersections provides a significantly better approximation for predicating pedestrian and bicyclist collisions.

Intervals compared	Individual intersections	5 adjacent intersections	Signalized vs. un-signalized
1 YR	3.19		
5 YR	1.55	0.69	1.05

Based on our analysis, we recommend the following for traffic engineers and planners:

1. Before starting a safety evaluation, develop a comprehensive database of the area to be considered. The web-based query system of statewide collision data for California developed by SafeTREC (http://www.tims.berkeley.edu/) can be used to help build a collision database.
2. Conduct a descriptive analysis of collision numbers and clusters, followed by an evaluation of sites at different scales with the intent of increasing the stability of estimates of expected injuries through methods described above.
3. Conduct an evaluation of the benefit-cost of different strategies. The tool developed by SafeTREC for the HSIP program is available for this purpose.

Further research is needed to refine and extend these data systems and methods:

1. Develop methods of analyses based on Bayesian techniques to estimate expected injuries. This will require constructing databases with infrastructure inventory data and reliable estimates of vehicle and pedestrian volume.
2. Identify specific features of the infrastructure associated with pedestrian and bicyclist risk, thereby permitting strategies that focus on sites with these specific features and increase the benefits of Bayesian analysis.
3. Identify techniques which will allow to cluster intersections by multiple features.
4. Explore the implications and benefit-cost of strategies above the specific site-level, specifically, those involving extended street segments, clusters of intersections, or systemic approaches.
5. Develop tools for evaluating the impact of pedestrian and bicyclist injury countermeasures on Level of Service (LOS) for vehicle traffic. This is important for being able to provide optimal level of service for all modes of traffic.

Strategies for implementation include: (i) provide training for traffic engineers and planners, (ii) design Internet-based or stand-alone tools that incorporate methods developed in this study as well other available methods; (iii) conduct applied research to extend and refine the methods and approaches developed in this project, and, finally, (iv) propose that methods and approaches for identifying sites with optimal potential be incorporated as one of the goals of the Strategic Highway Safety Plan (SHSP) process.

Table of Contents

List of Figures

List of Tables

Introduction

Current roadway safety data systems and methods of analysis have focused on motor vehicle and transit performance and efficiency. However, as more attention is paid to reducing energy use and associated vehicle travel and emissions, walking and bicycling are emerging as important modes in what promises to be an increasingly multi-modal system. Parallel data systems and methods of analysis for pedestrians and bicyclists are crucial if we are to fully incorporate them into this multi-modal transportation system.

A prime example is the San Pablo SMART Corridor, a system of freeways and major arterials that serves the east shore of the San Francisco Bay, running through five cities, from downtown Oakland north to the City of Hercules. It uses intelligent transportation system (ITS) technologies (such as video monitoring of intersections), signal timing, and the integration of a rapid bus system to increase and enhance transportation mobility within and throughout several East Bay communities.

In this study, funded by the California Department of Transportation (Caltrans), methods are developed to support multi-modal transportation, with a focus on methods for identifying sites where there is a potential for significant reductions in pedestrian and bicyclist injury.

San Pablo Avenue, the main arterial in the SMART Corridor, is a prominent feature in the East Bay. As such, it has been the subject of several projects conducted by SafeTREC in collaboration with other research groups. They include:

- A detailed report on a section of the SMART Corridor within the City of Berkeley (Pedestrian and Bicycle Safety Evaluation in a SMART Corridor, Berkeley Segment, May 2006)
- San Pablo Avenue Pedestrian Signal Timing Optimization (August 2006, presented at the Transportation Research Board)
- Pedestrian and Bicycle Safety Evaluation in a SMART Corridor, Berkeley, California (August 2006, presented at the Transportation Research Board).

A striking observation from the studies we have conducted of San Pablo Avenue and numerous other locations is that pedestrian and bicyclist events are spatially clustered to a large degree, meaning that they tend to be denser in some locations than in others. Variations in pedestrian and bicyclist volume may be a partial explanation: more pedestrians and bicyclists increase the chances of pedestrian and bicycle crashes. Variations in characteristics of the infrastructure may also contribute to clustering. For example, the length of pedestrian signal cycles varies dramatically among the five different jurisdictions that govern San Pablo Avenue.[1] The same is found for other pedestrian and bicyclist safety infrastructure elements.

Along with spatial clustering, there is considerable variation over time. This means that assuming that no major safety changes were implemented during the study period, there may be dramatic changes in the number of events from one year to the next in a certain area. This is more pronounced at individual sites (e.g., an intersection) because the number of events may be so small as to be statistically invalid, making it difficult to accurately predict where a collision will occur. This in turn makes it difficult to direct resources where they are most likely to prevent collisions.

Improving "Expected Number of Injuries" to Better Achieve "More Bang for the Buck"

The Federal Highway Administration (FHWA) has an explicit policy that "safety funds are to be used on the most effective treatments at the locations with the greatest needs, and that we are to use the best available data to determine the best treatments for each State's needs" and that this involves "leveraging funds and focusing spending on projects where the return on investment will be maximized."[2] Ezra Hauer, a noted researcher in this area, argues that the goal is to design programs so as to prevent the largest number of fatalities/injuries per dollar spent, i.e., the "most-bang-for-the-buck [MBB] principle." Hauer writes: "The MBB principle has controversial implications. And yet, to balk at implications of the MBB principle means that it is justified to save one accident when, for the same money, more than one could be saved. Such justifications are not easy to find. Therefore, in this paper, the guidance of the MBB principle is heeded."[3]

The question then, is how do we achieve the goal of maximizing return on investment in the area of pedestrian/bicyclist safety? In preventing pedestrian/bicyclist injury, it is generally impossible to treat an entire jurisdiction, since resources are always limited. Therefore, any program or plan to reduce pedestrian/bicyclist injury must begin with an assessment of where to direct resources.

The MBB principle is expressed in terms of benefit-cost calculations, i.e., Benefit-Costs Ratio = Savings of a treatment / Cost of a treatment. The "savings" is really *expected* savings since we don't know what the *actual* savings will be. The expected savings is expressed in terms of the number of injuries (or other events) prevented multiplied by the dollar value associated with those events.

---

[1] Nguyen A, Ragland DR. San Pablo Avenue Signal Timing Optimization. Safe Transportation Research and Education Center (SafeTREC). August 1, 2006. Available at: http://tinyurl.com/3r7exsv (accessed on June 1, 2011).

[2] Lindley JA. Achieving Maximum Results in Safety. Memorandum. Federal Highway Administration, U.S. Department of Transportation, May 17, 2006. Available at: http://safety.fhwa.dot.gov/hsip/policy_guide/memo051706.cfm (accessed May 28, 2011).

[3] Hauer E, Kononov J, Allery B, Griffith MS. Screening the Road Network for Sites with Promise. Transportation Research Record 1784. Paper No. 02-2182.

The number of events we *expect* to prevent is derived by multiplying the number of expected events by the expected reduction in injuries, i.e., the Collision Reduction Factor (CRF), where the CRF is the expected percent reduction in collisions.

If we assume for the present purposes a fixed dollar amount associated with an event, a known CRF for a particular countermeasure, and a cost for the countermeasure, then the question becomes one of finding an estimate of the number of events expected if nothing is done, i.e., if the countermeasure is not installed. This section, and most of this report, focuses on that question.

Until recently, the predominant approach for identifying high-risk collision locations (whether vehicle-only collisions or collisions involving pedestrians and bicyclists) was a two-part process that (1) focused on individual locations (e.g., particular intersections or crossings) and (2) used past history to generate an expected number of collisions.

However, both steps in this process are subject to phenomena that undermine their ability to effectively identify high-risk collision sites. In the case of Step 1—focusing on an individual location—typically, the frequency of pedestrian/bicycle events is fairly low. This leads to general instability in estimates of "expected injuries," which results in increases in both false positives (identifying a site as high-risk when in fact it is not) and false negatives (failing to identify sites that have a large number of events). Step 2—using past history to calculate future collisions—is affected by the phenomenon of "regression to the mean (RTM)." On one hand, RTM means that sites that were high in a certain time period will generally be lower in another time period, i.e., the injury frequencies will "regress to the mean." Treatment of such sites will yield less benefit in the future. On the other hand, RTM also means that sites with a relatively low number of collisions relative to the rest are also likely to be closer to the mean, i.e., to have more collisions in the future than would be predicted based on past history.

Consider a situation where calculations use historical data (say over one or more years) in a set of sites to estimate the "expected" number of collisions in the future, for the purpose of calculating the expected benefit-cost ratio for an intervention. From an intuitive viewpoint, there are two striking phenomena. First, a cluster of events (say, two or three) might appear at a location where there had been none for several years previously. In this case those sites would *not* have been treated—based on their past history of no events—and an opportunity to prevent those events would have been lost. Second, a cluster might appear at a particular site one year, followed by several years with no collisions. In this case the resources assigned to that site based on the cluster that occurred in that one year would have prevented fewer events than expected based on this immediate history, and the same resources would have yielded a higher "return on investment" at another site.[4]

[4] This phenomenon has some striking implications in areas where the occurrence of an event is relatively rare at the level of an individual site. In one case, the death of a young boy occurred at a location that had not had an injury collision in the span of our database (10 years). Based on consideration of past history

Of course, one option is simply to treat every site. This would address every event that could occur. However, resources are always limited, and so we must consider approaches to improve efficiency of selecting sites. In several studies we have explored methods for identifying larger areas in which risk may be elevated due to systemic issues (e.g., high vehicle speeds, poor signage, etc.). In a separate project for the San Pablo SMART Corridor we identify a number of clusters of intersections in which risk was distinctly higher than for other locations on the Corridor using a method called "Zone Analysis."[5,6] Previously, we conducted a similar study showing clusters of pedestrian collisions in the City of San Francisco.[7] In a project to identify the top 5 percent of highway risk locations in the State of California, we have explored several methods for identifying high-risk segments (e.g., ½ mile in length) for the non-State Highway System (SHS) or roadways.[8] We have noted that high-risk segments themselves are often clustered, perhaps even adjacent to one another, suggesting that individual segments form part of a large area, or corridor, where risk is elevated. SafetyAnalyst, an approach developed by FHWA to identify high-risk roadway locations, suggests that methods to identify corridors be developed (as opposed to individual locations), but no systematic approach is suggested.[9] Finally, we have reviewed the literature to determine some of the advantages and disadvantages of choosing locations for treatment at different scales (Appendix 3).

Using data from San Pablo Avenue for 10 years (1998-2007) we examine several different approaches to counteract these phenomena and generate more accurate recommendations for treating sites in the corridor. To do so, we examine and compare four approaches to developing statistically more stable estimates and discuss their characteristics and their application in relation to the traditional two-step approach of choosing individual sites based on their past history. We also demonstrate the rate of false positives and negatives under the traditional two-step approach, as well as gains and losses in statistical stability associated with the four alternative approaches under consideration.

The four approaches are:

---

this site would not have merited treatment. (SafeTREC UC Berkeley Campus Periphery Safety Project, 2010, forthcoming).

[5] Ragland DR, O'Connor TO. Pedestrian and Bicycle Safety Evaluation in a SMART Corridor. E-Scholarship, UCB-ITS-PRR-2008-25.

[6] Zone Guide for Pedestrian Safety. U.S. Department of Transportation, National Highway Traffic Safety Administration, Federal Highway Administration. 1998 DOT HS 808 742.

[7] San Francisco PedSafe Phase II. Final Implementation Report and Executive Summary. Prepared for the U.S. Department of Transportation, Federal Highway Administration by Cooperative Agreement DTFH61-02-X-00017. Available at: http://safety.fhwa.dot.gov/ped_bike/tools_solve/ped_scdproj/sf/pedsafety_sf.pdf.

[8] Five Percent Report for California, FHWA, 2010. Available at: http://safety.fhwa.dot.gov/hsip/fivepercent/2010/index.cfm?state=ca (accessed May 30, 2011).

[9] Safety Analyst. AASHTO. Available at: http://www.safetyanalyst.org/tools.htm (accessed May 31, 2011).

- Increasing the Time Horizon for Events, either Years of History or Post-Treatment Follow-up
- Increasing the Geographic Scale (from specific sites to corridors, zones, or an entire network)
- Combining Sites with Similar Characteristics
- Creating Estimates Using the Bayesian Method

Table 1. Description and Comparison of Strengths and Weaknesses of Different Approaches to Site Selection

Approach	Description	Strengths	Weaknesses	Comment
A. Choose Specific Sites Using Past History	Calculate BC for individual sites and rank.	Intuitive, methods exist to identify sites.	Instability of estimates of expected injuries, especially if injury rates are low.	Traditional approach followed by many current jurisdictions and funding programs.
B. Increasing Time Horizon for Events, either Years of History and/or Follow-up	Same as "A" but increase years of history and/or years of follow-up.	Gain numbers and therefore increase stability of estimates of expected injuries.	Potential bias if changes take place over time (i.e., greater chance of change with increasing time).	Very effective in increasing stability of estimates if no reason to suspect historical change in conditions.
C. Increasing Geographic Scale (from specific sites to corridors, zones, or entire network)	Same as "A" but increase scale of "sites" in order to increase numbers.	Gain numbers and therefore increase stability of estimates of expected injuries.	Need to spread countermeasures over a greater area or number of sites.	Very effective if treatment costs per unit of area or number of sites can be kept low.
D. Combining Sites with Similar Characteristics	For example, combine mid-block crossings.	Gain numbers and therefore increase stability of estimates of expected injuries. The same countermeasures installed at all locations, possible economy of scale.	Need to spread countermeasures over a greater distance or number of sites.	Very effective if treatment costs per unit of area or number of sites can be kept low and there can be an advantage of consolidating engineering analyses.
E. Creating Estimates Using the Bayesian Method	Create model of the network and apply combined modeled estimate of injuries with history of injuries.	Increase stability of estimates of expected injuries.	Need for network database with relevant variables.	Can be combined with any of the above if data is available.

Choosing Potential Countermeasures

Until fairly recently there had not been attention paid to choosing appropriate countermeasures once a site had been selected. While it is not the purpose of this report to cover this topic in detail, it is a necessary step in the process of developing a plan for reducing pedestrian and bicycle fatalities and will be considered briefly here. FHWA has commissioned several Internet-based resources for assisting in identifying the main issues at a particular site and selecting the appropriate potential countermeasures. Two sites in particular are linked to the University of North Carolina (UNC) Pedestrian and Bicycle Information Center. These two sites, one for walking and one for biking, are: http://www.walkinginfo.org/pedsafe/ and http://www.bicyclinginfo.org/bikesafe/. This step is necessary for the identification of countermeasures for comparison in benefit-cost calculations.

Framework for Conducting Benefit-Cost (BC) Analyses

We have listed in Table 1 several approaches to improving stability of the expected injury rate. We now need a method for calculating benefit-cost (BC) that has at least the following specifications: (i) compares across different levels of scale; (ii) considers different levels of injury; and (iii) compares different types of injury (e.g., pedestrian, bicyclist, motor vehicle occupant).

Within the context of this study we have developed an approach that uses Crash Reduction Factors (CRFs) from the FHWA Desktop Reference for Crash Reduction Factors[10] and the FHWA Crash Modification Clearinghouse[11,12] that has the following characteristics: First, the method is capable of comparing countermeasures that have been applied on different scales, e.g. a countermeasure at an individual site could be compared with a countermeasure at a corridor or systemic level. Second, treatment options applied mainly for reducing pedestrian or bicyclist collisions can be compared with treatment options targeting mainly vehicle-vehicle collisions. Since for many countermeasures the FHWA Guidebook specifies the type of collision that is reduced by the countermeasure, the approach can be used to apply CRFs differentially to the various collisions occurring at a site, a set of sets, a corridor, or a zone.

[10] Desktop Reference for Crash Reduction Factors. Federal Highway Administration, Office of Safety. Available at: http://safety.fhwa.dot.gov/tools/crf/resources/fhwasa08011/ (accessed on May 28, 2011)

[11]Crash Modification Factor Clearinghouse. Federal Highway Administration. Available at: http://safety.fhwa.dot.gov/tools/crf/resources/(accessed on May 28, 2011)

[12]Crash Reduction Factors and Crash Modification Factors are algebraically equivalent. Available at: http://safety.fhwa.dot.gov/tools/crf/(accessed June 1, 2011).

In the context of another project we have developed this approach to create a stand-alone tool that is now being used by Caltrans for the HSIP application process for local (i.e., non-State Highway) jurisdictions.[13,14]

Once sites are identified for study, the information needed is the set of potential countermeasures and their estimated costs and the expected collisions at the site (or corridor or area) under investigation. The tool uses standard formulas for benefit-cost calculation, which are included in the Highway Safety Improvement Plan (HSIP) program guide [15] (Appendix 6). The tool also links to extensive HSIP safety resources.[16]

---

[13] HSIP Program Guidelines and Application Tool HSIP Program Guidelines and Application Tool, Caltrans Division of Local Technical Assistance. Available at http://www.dot.ca.gov/hq/LocalPrograms/HSIP/apply_now.htm and http://www.dot.ca.gov/hq/LocalPrograms/HSIP/tool_instructions.htm (accessed on May 31, 2011).

[14] Note. The tool must be "enabled" to be viewed.

[15] Chapter 9: Highway Safety Improvement Program Guidelines. Cycle 4 – 2010-2011 Federal Fiscal Year. Highway Safety Improvement Program (HSIP). Local Assistance Program Guidelines. Available at: http://www.dot.ca.gov/hq/LocalPrograms/HSIP/Documents/HSIP_Guidelines.pdf (accessed May 30, 2011).

[16] http://safety.fhwa.dot.gov/hsip/ (accessed May 30, 2011).

Study Area and Data Preparation

The study area, shown in Figure 1 is a 16.5-mile section of San Pablo Avenue (SR 123), an arterial corridor in San Francisco's East Bay. Some features are listed below:

- Runs from Frank H. Ogawa Plaza (red arrow on right) in downtown Oakland to Solano Avenue in Richmond (red arrow on left).
- Passes through 5 different cities: Oakland, Berkeley, Albany, El Cerrito, and Richmond.
- Crosses 180 intersections that are on average approximately 484 feet (0.09 miles) apart.
- Intersects with six major Vulnerable Road User (VRU) arterials—arterials with high levels of pedestrians and bicyclists: 17th Street and 40th/Adeline Streets (Oakland), Ashby Avenue (Oakland-Berkeley border), University Avenue (Berkeley), Solano Avenue and Hill St. (El Cerrito) (labeled in Figure 1).
- Environment along this corridor varies significantly as it moves through different cities and land-use characteristics, e.g., mega-retailers at Hill Street in Richmond, local retail at Solano Avenue in El Cerrito, at University Avenue in Berkeley, at Ashby and at 40th/Adeline Streets in Oakland, and a compact downtown business district at 17th Street in Oakland. (Appendices
- Appendix 1. Photographs of VRU arterials along San Pablo Avenue) shows photos of the intersections with the major VRU arterials.
- Was the site of 413 VRU (pedestrian and bicyclist) collisions between 1998 and 2007, which corresponds to an annual collision rate of about 0.23 collisions per intersection. Circles indicate where these took place over the 10 years studied.

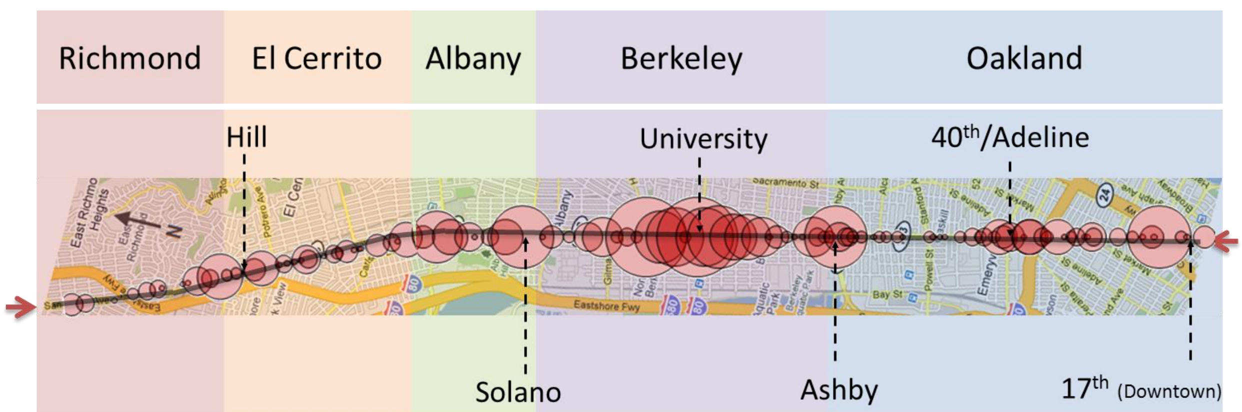


Figure 1. Cities, VRU arterials, and VRU crashes in the San Pablo Avenue Study area

A database including all VRU collisions was constructed using the Statewide Integrated Traffic Records System (SWITRS) maintained by the California Highway Patrol (CHP).

The database includes all pedestrian and bicyclist collisions from 1998 to 2007. Each collision was geocoded and matched to the closest intersection. The running distance from the first intersection through the last one was also calculated. In total 413 pedestrian and bicyclist collisions occurred between 1998 and 2007, corresponding to an annual collision rate of about 0.23 collisions per intersection.

The red circles in Figure 1 represent the relative number of collisions for each intersection, with larger circles indicating higher numbers of crashes (the actual number of collisions are included in Appendix 2).

This figure illustrates (1) that there is a high concentration of collisions at some intersections and (2) that there is a high concentration within groups of intersections, suggesting there are factors affecting not just the individual intersection, but affecting an entire set of intersections. The first observation would suggest that resources be directed at individual intersections with the highest number of collisions. The second observation would suggest that there is a value in extending this approach to clusters of intersections that have a high number of collisions. In the following sections, we examine these and other approaches for their usefulness in identifying methods for choosing treatments for VRU safety.

Analysis at the Intersection Level Using a Single Year

This section evaluates the strengths and weaknesses of studying individual intersections using a single year as the sampling period. Table 2 summarizes the distribution of the frequency of VRU collisions per intersection for each of the 10 years, ranging from the percent of intersections with zero (0) collisions to the percent of intersections which had four (4), the greatest number of VRU collisions at any one intersection in this time period.

On average, 82.4 percent of the intersections in the study area did not have a single VRU collision in an individual year; this means that in each individual year 100 percent of the collisions take place at approximately 20 percent of the intersections. Because the vast majority of individual intersections did not have any collisions, this complicates estimating the expected number of injuries and demonstrates the challenges of studying VRU crashes.

As will be illustrated in the next section, a large number of the 80 percent of intersections with no collisions in a particular year in fact did have collisions in subsequent years.

Table 2. Distribution of VRU collision frequency per intersection in San Pablo Avenue study area by year

	% of Number of VRU Collisions per Intersection (from zero to 4)					
	0 VRU Collisions	**1 VRU Collision**	**2 VRU Collisions**	**3 VRU[17] Collisions**	**4VRU[17]Collisions**	**Total VRU Collisions**
2007	78.9%	15.6%	3.9%	1.1%	0.6%	52
2006	83.9%	11.7%	3.9%	0.6%	-	38
2005	77.8%	17.2%	3.3%	1.1%	0.6%	53
2004	86.1%	12.8%	1.1%	-	-	27
2003	83.3%	12.8%	2.8%	0.6%	0.6%	40
2002	82.8%	13.9%	2.2%	0.6%	0.6%	40
2001	83.3%	11.1%	4.4%	1.1%	-	42
2000	79.4%	13.9%	6.1%	-	0.6%	51
1999	86.7%	11.1%	1.1%	1.1%	-	30
1998	82.2%	14.4%	2.2%	1.1%	-	40
Average	82.4%	13.5%	3.1%	0.7%	0.3%	41.3

Table 2 also shows the annual *total* of collisions for each year. It is important to note the variation in the total per year, ranging from a high of 53 in 2005 to a low of 27 the year

[17] The actual locations of these crashes are available in Appendix 2

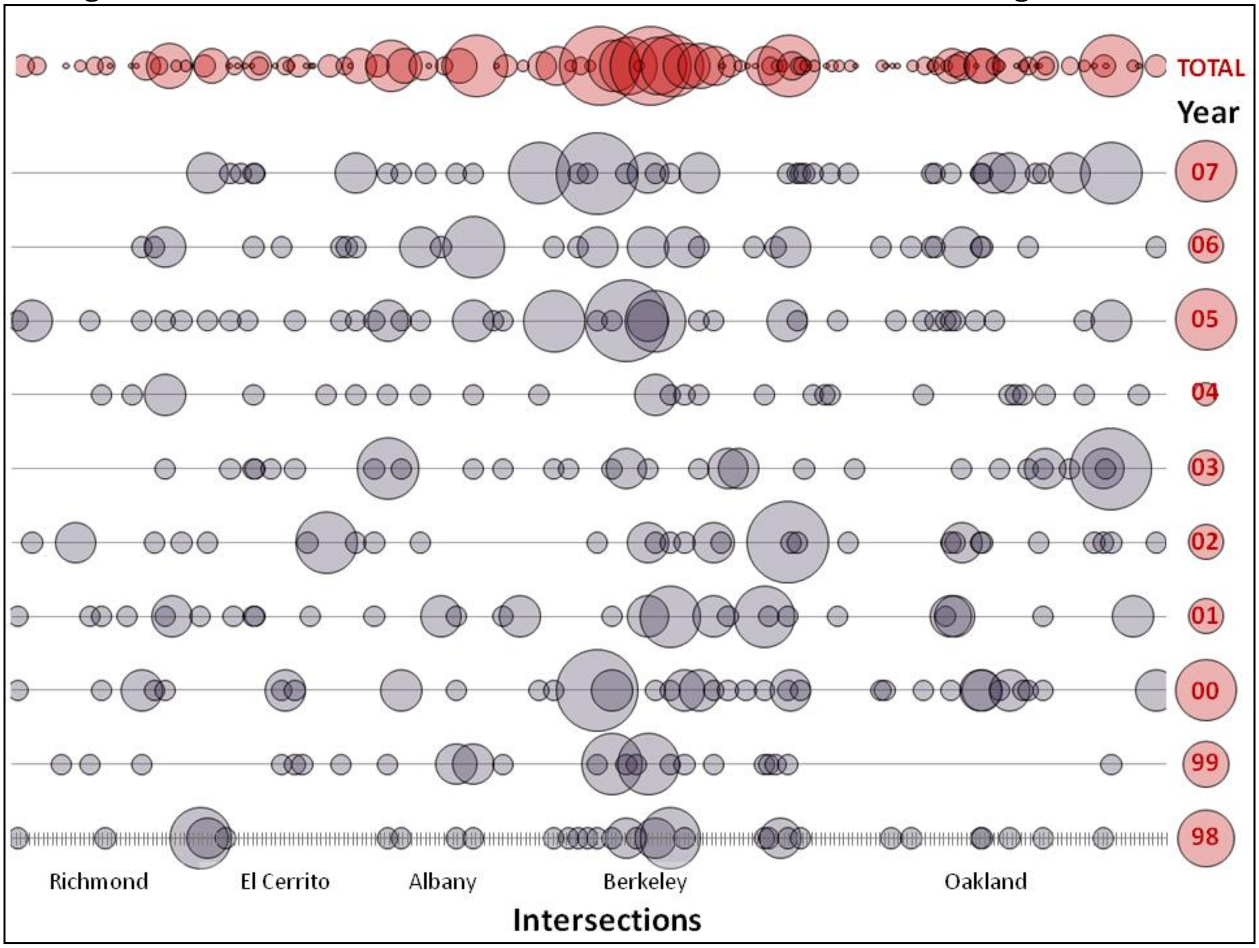
before, in 2004. It is striking that this variation by year occurs despite the fact that these numbers are aggregated over a large number—180—intersections. It is difficult to determine whether this represents an actual change in crash risk or is merely a statistical anomaly. It is important to note that the insights identified here assume that no major safety changes were applied to the study area during the study period.

Visualization of the Data

To help understand the patterns of collisions over space and time we developed a tool that illustrates both patterns simultaneously. Figure 2 shows this visualization. The horizontal axis represents the running distance along the corridor from Solano Avenue in San Pablo (left) to 17th Street in Oakland (right). Since the collisions were matched to the closest intersection, the annual collisions for each specific intersection are presented along a perpendicular line. The vertical axis represents the different years. The size of the gray circles along each year line represents the annual number of collisions at each of the intersections (the largest circle is four collisions); the red circles on the right of each year line are the total annual collisions across all intersections (on a different scale); the red circles on the top are the total number of crashes in each individual intersection over the entire study period. Using this diagram we can look into the collision counts and locations in more detail.

An examination of Figure 2 reveals insights into crashes on San Pablo Avenue in particular and along a corridor in general. First, we can see that clusters that appear stable for a combined 10-year period actually show very significant variation over time.

Figure 2. Annual number of collisions for individual intersections along SPA



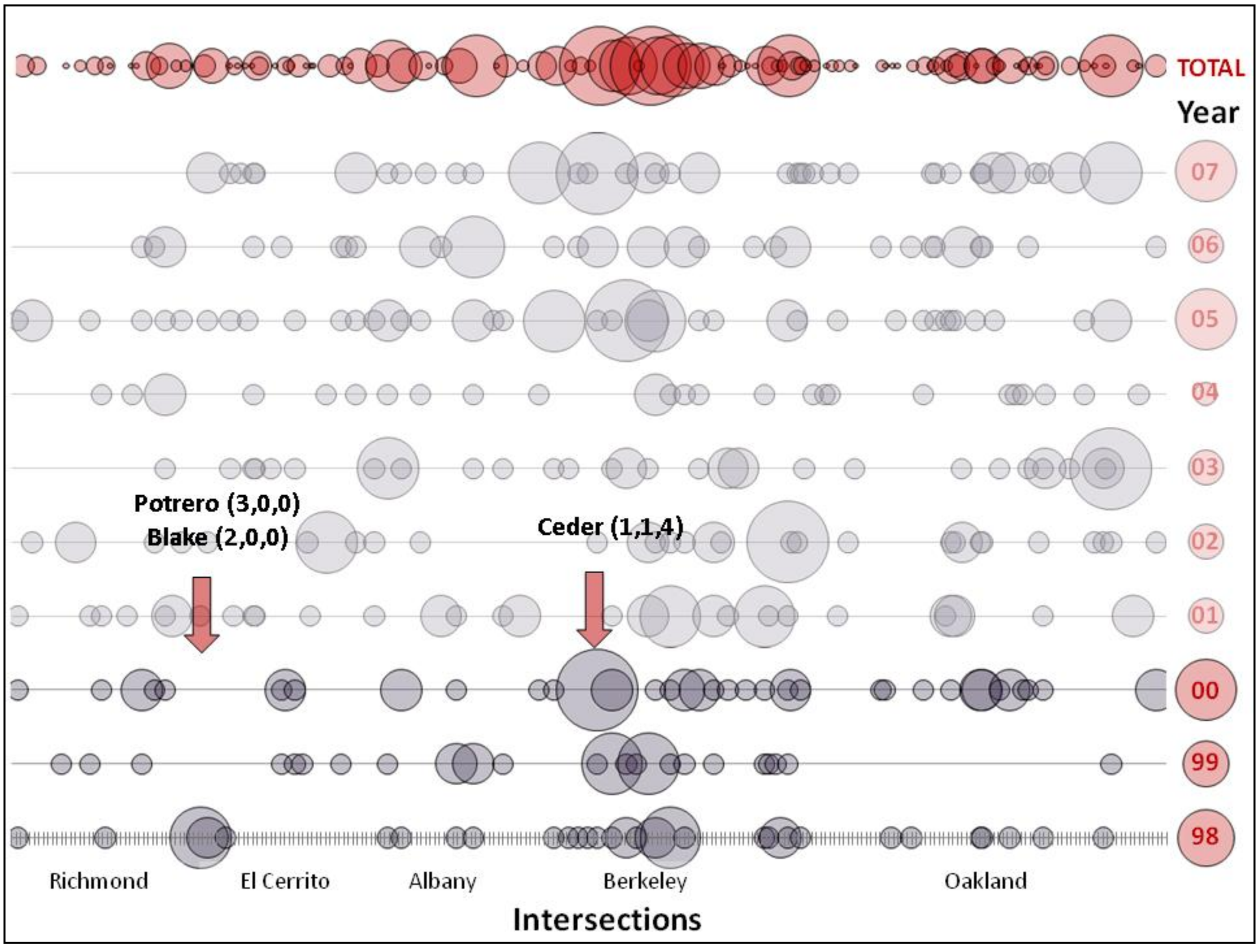
There are a number of instances where a cluster of collisions will be observed and then not appear for some time. For example, at the intersections with Potrero Avenue and Blake Street, shown in Figure 3, there were three and two collisions, respectively, in 1998, followed by several years with no collisions. It could be argued that expending resources at these two sites based on the 1998 collision data would have accomplished little because no collisions were going to occur there the next several years into the future.

On the other hand, there are instances where a major concentration of collisions may occur when there had been a relatively low number for some time. In the year 2000 four collisions occurred at Cedar Street, preceded by only one each year for the previous two years.

Considering these two examples, if a decision had been made to treat Potrero Avenue, based on the 1998 data, and not treat Cedar Street, there would have been a failure to prevent the collisions at Cedar Street; resources devoted to Potrero Avenue would have been expended needlessly.

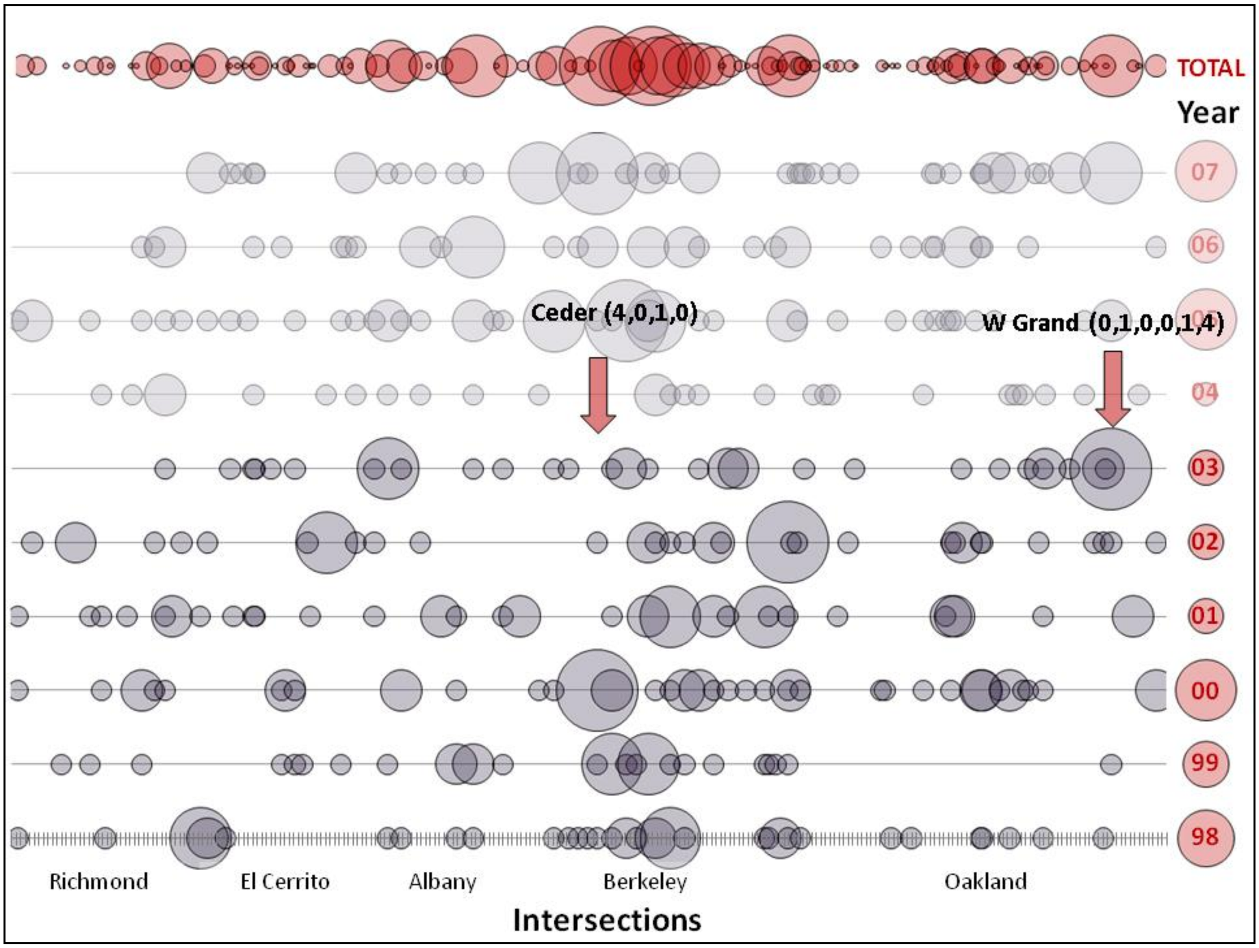
Of course, if a decision had to have been made in 1998, it would have been made without the knowledge of the future number of collisions at either intersection. In the subsequent sections we will evaluate whether a better decision could have been made in 1998 based on information that was available at the time.

Figure 3. Annual number of collisions for individual intersections along SPA and at Portero Ave., Blake St. and Cedar St.



Another example is West Grand Avenue, a major arterial crossing San Pablo. As shown in Figure 4, four collisions occurred at the intersection in 2003. However, the history prior to 2003 would not have indicated that resources would have prevented collisions, i.e., the expected number of collisions based simply on past history would not have identified this location as a site with a cluster of four collisions occurring in 2003. Again, the question is whether a better decision could have been made based on information available prior to 2003.

Figure 4. Annual number of collisions for individual intersections along SPA and at Cedar St. and West Grand Ave.



A Measure of Fit between Two Data Periods

We can estimate the differences between two years by using the Root Mean Square Error (RMSE). This statistic determines how well two datasets conform using the residuals. The estimator for the RMSE is:

Where T_1 and T_2 are vectors containing the number of crashes for intersection *i* in the corresponding data set, $t_{1,i}$, $t_{2,i}$. Since we are interested in making comparisons between different alternative approaches we need to normalize by the mean of the observed values to account for the different scales, which is the Coefficient of Variation (*CV*) of the *RMSE*, *CV*(*RMSE*). Table 3 shows the *CV*(*RMSE*) values for pairs of consecutive years. Lower values represent a better fit, and in general values of *CV* that are less than 1 are considered low variance. The average *CV(RMSE)* is 3.19, which is considerably higher than 1, and will be compared to the values we will obtain in the subsequent sections.

Table 3. *CV*(*RMSE*) of the number of collisions for consecutive years

	Base year								
	1998	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**
1999	3.16								
2000		3.09							
2001			3.12						
2002				3.21					
2003					3.55				
2004						3.56			
2005							3.11		
2006								2.93	
2007									2.95

We have established that basing decisions on individual intersections and single years is of limited efficacy and will yield substantial numbers of two types of errors: (i) false positives, i.e., selecting a particular intersection for treatment when in fact there would be a fairly small number of collisions there in subsequent years and (ii) false negatives, i.e., failing to select an intersection when in fact that intersection would have a significant number of collisions in ensuing years. If, as is the case, the goal is to prevent the maximum number of collisions per dollar spent, prioritizing spending according to annual crashes at individual intersections would experience limited success.

Intersection Level Using Multiple Years

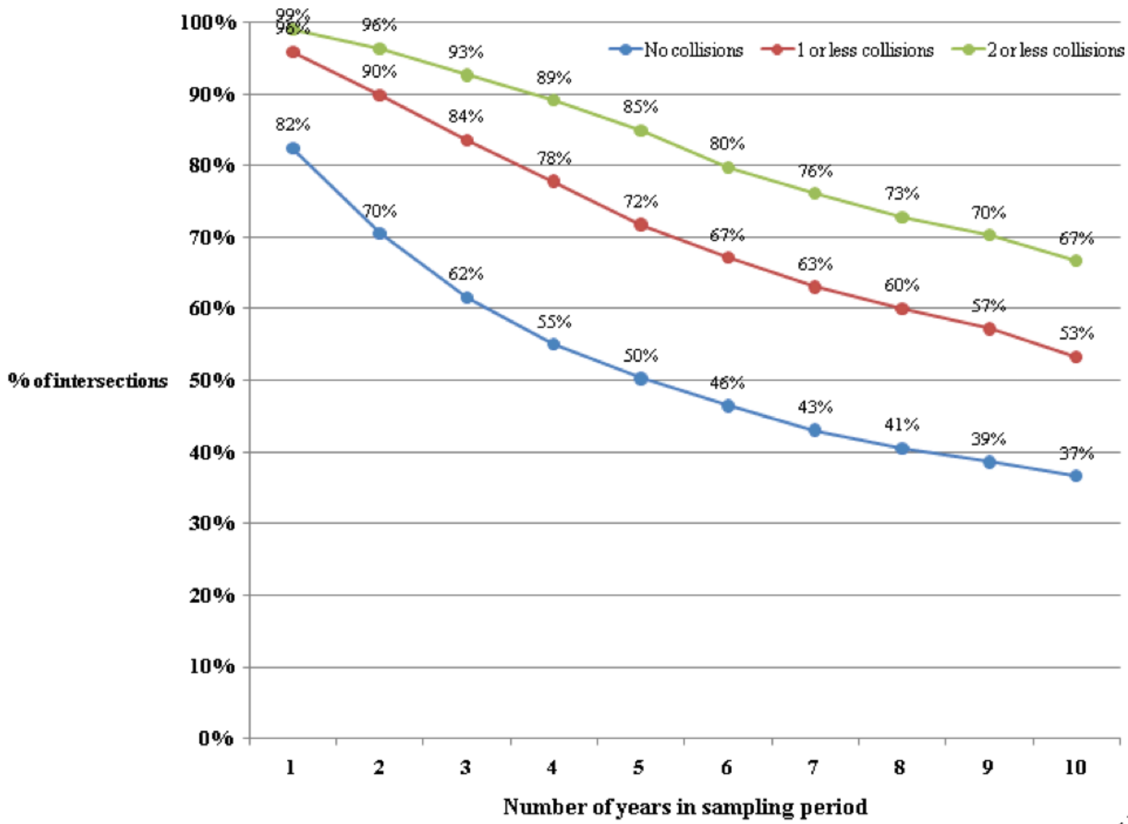
In this section we evaluate whether estimates for the expected number of collisions can be improved by increasing the sampling period beyond a single year. Table 4 shows the distribution of the number of crashes using sampling periods of different lengths, ranging from one to 10 years (the first row is the same as was shown in Table 2). As the sampling period increases, the percent of intersections without a crash drops from 82 percent (for a one-year sample) to 37 percent (for 10).

Table 4. Distribution of the number of crashes per intersection for 1998-2007

	Number of Collisions per Intersection												
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**14**
Years													
1	82%	13%	3%	0.7%	0.3%	0%	0%	0%	0%	0%	0%	0%	0%
2	70%	19%	7%	2.2%	1.0%	0.4%	0.1%	0%	0%	0%	0%	0%	0%
3	62%	22%	9%	4%	1.6%	1.2%	0.7%	0.1%	0%	0%	0%	0%	0%
4	55%	23%	11%	5%	2.1%	1.7%	1.3%	0.6%	0.1%	0%	0%	0%	0%
5	50%	21%	13%	7%	3%	1.6%	1.8%	1.4%	0.3%	0.2%	0%	0%	0%
6	46%	21%	13%	9%	4%	2.2%	2.1%	1.7%	0.8%	0.3%	0.1%	0%	0%
7	43%	20%	13%	9%	5%	3%	2.5%	2.1%	1.5%	0.3%	0.4%	0.1%	0%
8	41%	19%	13%	8%	7%	4%	2.0%	2.2%	1.9%	1.3%	0.7%	0.2%	0%
9	39%	19%	13%	8%	7%	6%	1.4%	1.4%	3%	0.8%	1.7%	0.3%	0.3%
10	37%	17%	13%	11%	6%	4%	4%	1.7%	1.7%	1.1%	1.1%	2.2%	1.1%

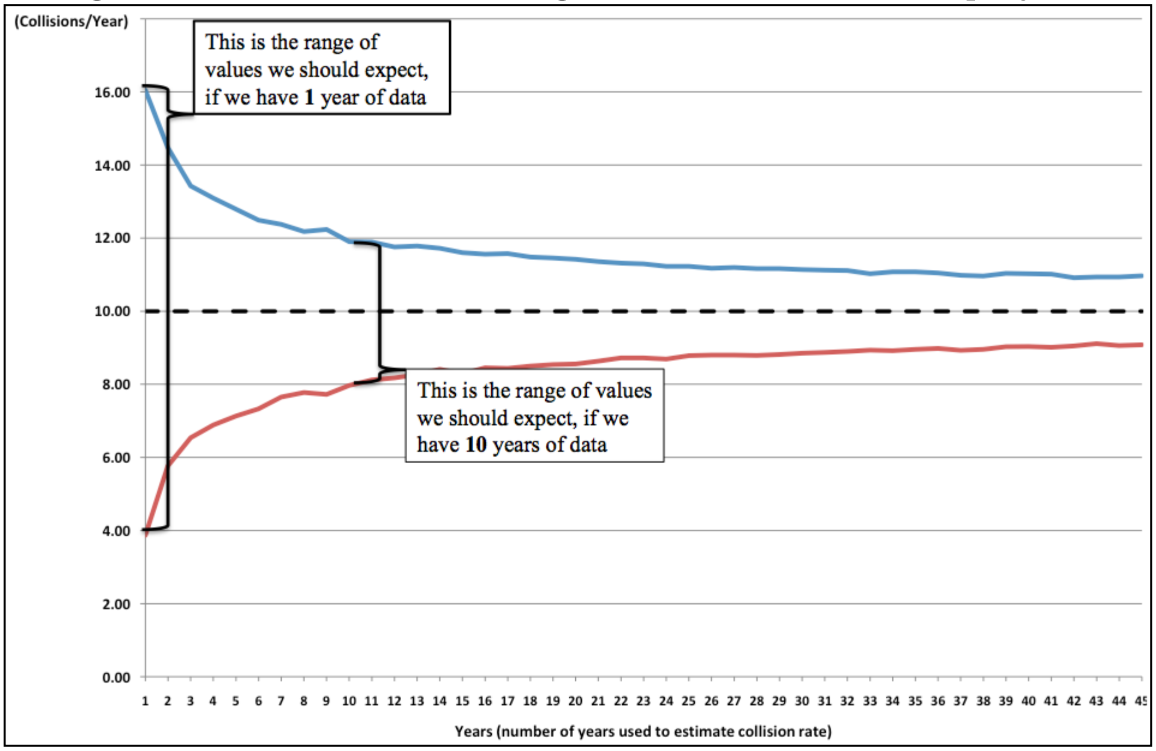
Figure 5 shows graphically how the distribution of the number of crashes changes as the sampling period increases. From bottom to top, the three lines represent three types of intersections: the bottom line represent intersections with zero collisions; the next line shows the percent of intersections with zero or one collisions; and the top line shows intersections with zero, one, or two collisions. This analysis shows that as the sampling period increases, predictability improves, the result of some of the noise in the data being "ironed out." However, there is a drawback to this approach: physical conditions may change over a long period of time, causing the grouping of intersections that are fundamentally different as time goes on.

Figure 5. Change in distribution of number of collisions per intersection for three different types of intersection—those with zero crashes, one or fewer crashes, and two or fewer crashes—over extended periods, in this case 10 years.



To further understand the benefit of increasing the sampling period we constructed a 95 percent Confidence Interval (CI) diagram for a single intersection, which is shown in Figure 6. First we chose an initial collision rate, in this case 10 per year per intersection (shown by the dashed line). We then constructed upper and lowers bounds assuming a Poisson distribution. We then followed the same procedure to construct the bounds around an average of two Poisson random variables and so on, up to an average of 45 Poisson random variables (for each of the 45 years). The horizontal axis represents the number of years used to estimate the CI's—running from one to 45. The diagram shows that the range of collisions expected to be observed when looking at a single year is approximately 4 to 16. Using data for two years, the range is narrower, approximately between 5.5 and 14.5. With 10 years of data, the range has shrunk still further: the values are between 8 and 10. (Note that when more than one year is used the values represent the average over the sampling period.) As more years are added the CI's become smaller, with a systematically shrinking confidence region around the initial 10-per-year collision rate.

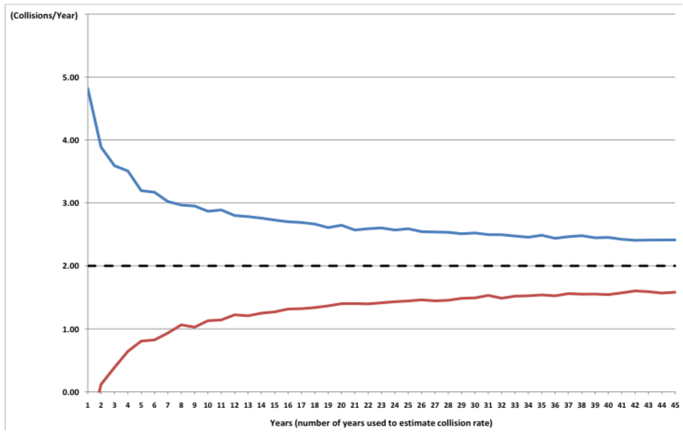
Figure 6. A Confidence Interval diagram for a collision rate of 10 per year



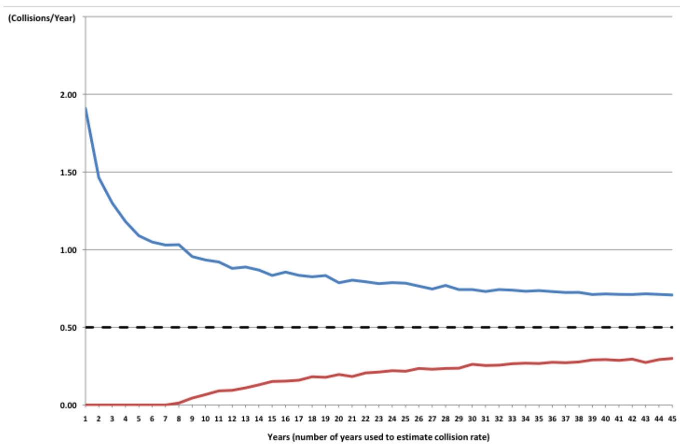
The diagram reveals that the CI's decrease most rapidly over the first three years, and then decrease more slowly until they converge. This may indicate that using a sampling period of three years, as is common, provides a good balance between reducing statistical variation and accounting for changes in the intersections over time.

However, the number of VRU crashes in the study area is 0.23 per year, orders of magnitude smaller than the 10-per-year used to build the diagram in Figure 6. We then constructed CI diagrams for collision risks whose magnitudes are closer to that of the study area: 2 per year, 0.5 per year, and 0.1 per year. The results are shown in Figure 7 (a-c).

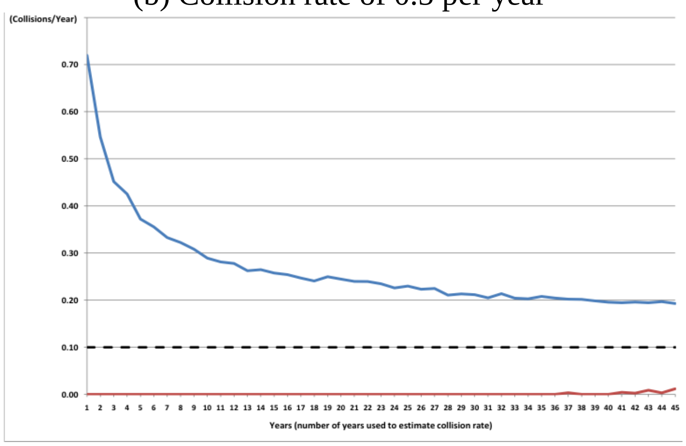
Figure 7. Confidence Interval diagrams for three lower collision rates (as indicated by dotted lines)



(a) Collision rate of 2 per year



(b) Collision rate of 0.5 per year



(c) Collision rate of 0.1 per year

The CI diagram in Figure 7 (c) shows that for the lowest collision risk the lower bound of the CI is truncated only barely rising above 0 in the final years of the sampling. This makes the area between the curves smaller, which reflects a reduction in the statistical variance, but it also reduces the rate of improvement as the sampling period is extended, since the lower bound remains at 0 for so long. However, the curves of the upper and lower bounds of the CIs shown in Figure 7 (b) and Figure 7 (a) demonstrate a convergence similar to the diagram in Figure 6; in this case, starting at year 5, suggesting that for low-crash intersections such as those in the study area, five years is the preferred sampling period to evaluate the benefits for estimating the number of crashes.

We can now use again *CV(RMSE)* to compare the fit of data for different sampling periods. Table 5 shows the *CV(RMSE)* values for periods between one and five years. The table shows that the overall trend is an improvement of *CV(RMSE)* as we increase the sampling period. Using five years of data had an *CV(RMSE)* of 1.55 compared to the one year periods which had an average of 3.19. This further supports the findings revealed using the CI diagrams.

Table 5. *CV*(*RMSE*) of the number of collisions for two periods

		Base period								
		1YR	**2YR**				**3YR**		**4YR**	**5YR**
		One year	**1998-1999**	**2000-2001**	**2002-2003**	**2004-2005**	**1998-2000**	**2001-2003**	**1998-2001**	**1998-2002**
1YR	**One year**	3.19								
2YR	**2000-2001**		2.01							
	2002-2003			2.25						
	2004-2005				1.97					
	2006-2007					2.01				
3YR	**2001-2003**						1.96			
	2004-2006							2.19		
4YR	**2002-2005**								1.74	
5YR	**2003-2007**									1.55

Analysis Using Five Years

We next performed an analysis to determine the value of 5-year sampling periods. The five years of crash data between 1998 and 2002 was designated the baseline, and the data for 2003 to 2007 was the follow-up. The comparison is shown in Figure 8. As expected, the 5-year sampling period made differences among the intersections easily visible, although though both periods show a similar cluster of crashes around the center of the study area.

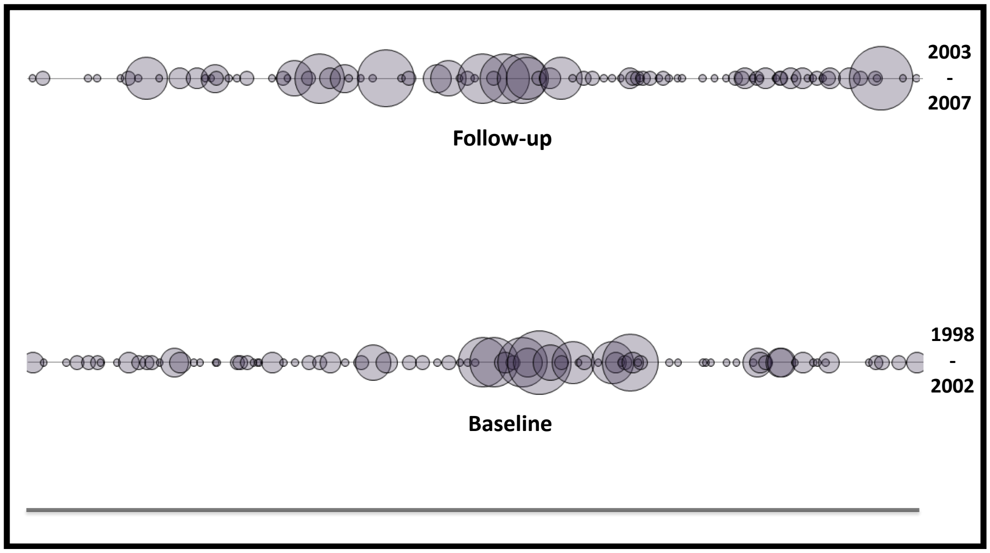


Figure 8. Number of crashes aggregated in two five-year samples: the baseline period (1998-2002) and the follow-up (2003-2007)

We then created a table showing the distributions of the number of collisions for each of the intersections during the baseline and follow-up periods (Table 6). It shows that about 50 percent of the intersections had zero crashes, and that the overall distributions for the two periods are comparable. However, this may be misleading because it does not indicate how the two periods compare at the level of individual intersections. We sorted the data according to crashes in the baseline period, as shown in Figure 9. Again, it shows that about 50 percent of the intersections had no collisions in 1998-2002; however we now see that those same intersections had 27 collisions in 2003-2007, which would not have been discernible using the baseline data alone.

Table 6. Distribution of the number of crashes per intersection for the baseline and follow-up periods

1998-2002			2003-2007		
Collision Count	**Number of Intersections**	**% of intersections**	**Collision Count**	**Number of Intersections**	**% of intersections**
0	93	51.7%	0	89	49.4%
1	36	20.0%	1	40	22.2%
2	24	13.3%	2	26	14.4%
3	13	7.2%	3	11	6.1%
4	5	2.8%	4	3	1.7%
5	2	1.1%	5	2	1.1%
6	2	1.1%	6	3	1.7%
7	3	1.7%	7	4	2.2%
8	1	0.6%	8	1	0.6%
9	1	0.6%	9	1	0.6%

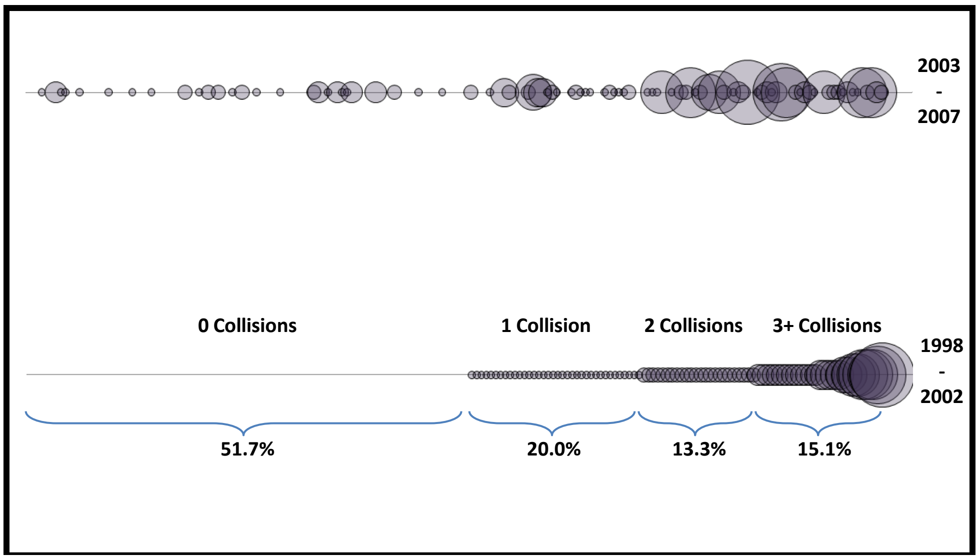


Figure 9. Number of crashes in the baseline and follow-up period sorted by the baseline

To further study how the baseline and the follow-up periods correlate we constructed cross tabulations of both periods (Figure 10). It shows that using the baseline period's zero-crash prediction was reliable for 66 intersections. These 66 are labeled as True Negatives (TN). However, for another 27 intersections, the zero-crash predictions were incorrect; these are labeled False Negative (FN).

		2003-2007 (follow-up)										
		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**TOT**
1998-2002 (baseline)	**0**	66	16	6	5							**93**
	1	14	11	7		3	1					**36**
	2	7	6	5	1		1	2	1		1	**24**
	3	1	4	3	3				1	1		**13**
	4	1		3				1				**5**
	5		1		1							**2**
	6		2									**2**
	7			1					2			**3**
	8				1							**1**
	9			1								**1**
	TOT	**89**	**40**	**26**	**11**	**3**	**2**	**3**	**4**	**1**	**1**	**180**

Figure 10. Cross tabulation of the number of crashes in the baseline and follow-up period

A more refined comparison can be done by restricting the comparison to intersections that have experienced a minimum number of collisions—in this case more than 2. The results of this comparison are shown in Figure 11. Only 27 intersections fit this criterion, and of them, only 10 had three or more crashes in the follow-up period. These are True Positives (TP), and 17 had less than three, and these are False Positives (FP). Of the intersections that had a low number of crashes in the baseline period, 138 still had a low crash rate in the follow-up period. These are True Negatives (TN), while 15 had a high crash rate. These are False Negatives (FN).

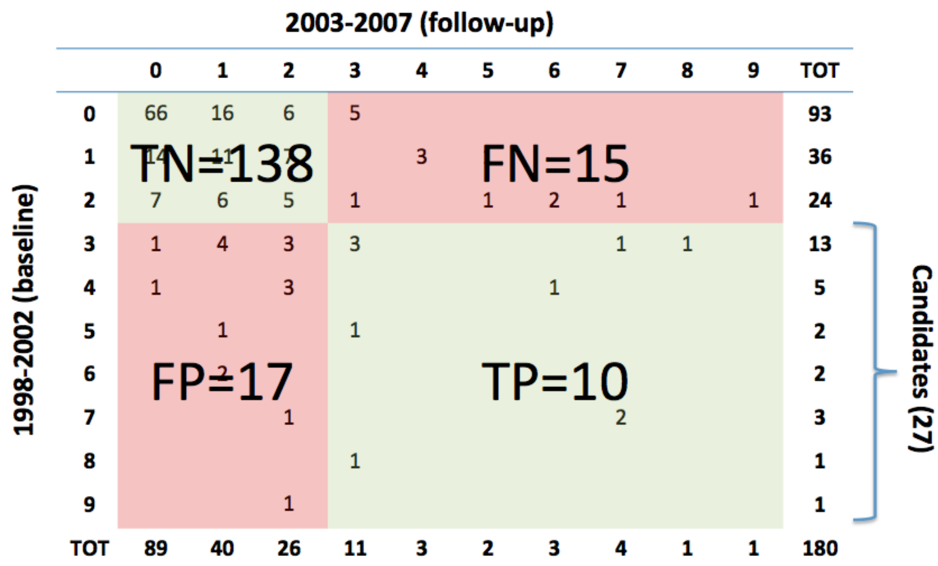


Figure 11. Number of candidate intersections, and the corresponding outcomes, for those with more than 2 collisions

False positives and false negatives have different implications. A false positive leads to expenditures that result in fewer savings than anticipated, and "opportunity costs" in the form of resources that might have been used to treat another intersection that would have produced greater returns on the investment. A false negative means missing a chance to make an expenditure that would have a high potential for preventing collisions, with the resources going to a less dangerous site.

Keeping in mind that the objective is to estimate collisions, it is also necessary to compare the number of baseline collisions that occurred in the candidate intersections to the number of follow-up collisions. Figure 12 shows three scatter-plots: (i) top left counts the intersections; top right counts the collisions in the follow-up period; and (iii) bottom left counts the collisions in the baseline period. It shows that the 27 candidate intersection were chosen based on 119 collisions. However, in the follow-up period there were only 73 collisions, of which 23 were FP. Of the 137 collision that occurred in non-candidate intersections, about half were FN (68). That means that about a third of the intersection would perform better than the threshold irrespective of any costly inspection and intervention. On the other hand about half of the intersections that were considered low collision frequency actually did exhibit collision rates that were higher than the threshold. Allocating resources based on this approach is inefficient and any successful reductions are difficult to link to interventions.

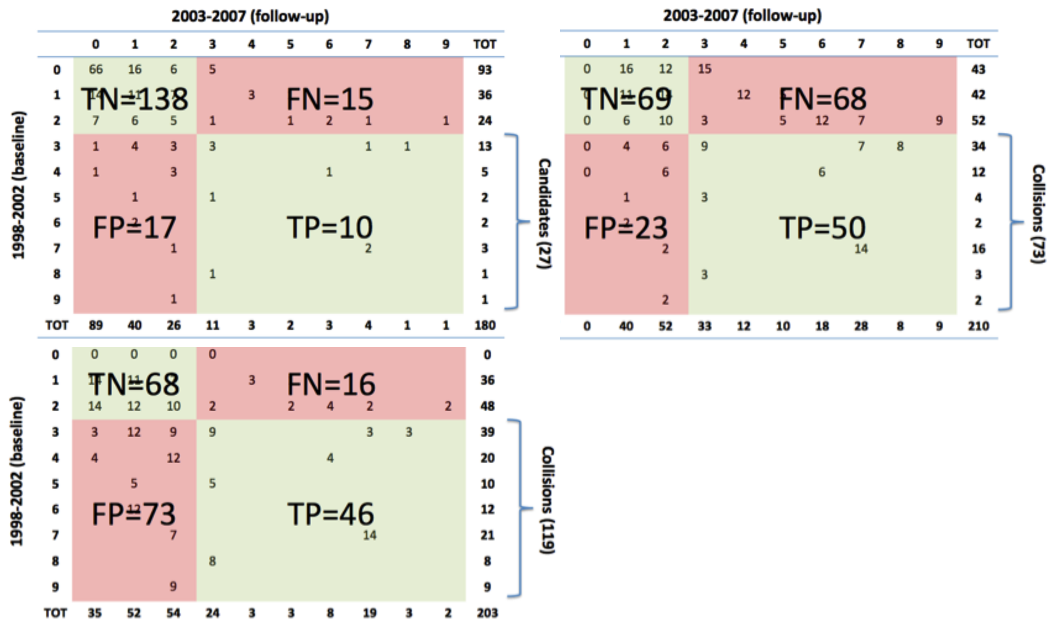


Figure 12. Number of candidate intersections, and the corresponding outcomes, for intersections, collision in the baseline period (bottom-left) and collisions in the follow-up period (top-right), for those with more than 2 collisions at the intersection level for a 5 year sampling period

For a greater understanding of the underlying mechanism resulting in false positives and false negatives, we reformatted the cross tabulation data in Figure 11 to a scatter plot (Figure 13). Note that the axes are flipped from Figure 11: the x-axis now reflects the number of collisions that occurred at an intersection in the baseline period, and the y-axis those in the follow-up period.

The size of the blue circles represents the total number of intersections that experienced that combination of collisions. For example, in Figure 11, the number "66" in the upper left corner represents the number of intersections that experienced zero collisions in each time period. That same group of intersections is represented in Figure 13 by the largest blue circle in the lower left of the plot.

The dark dotted diagonal line indicates the relationship if the same number of crashes occurred at an intersection in both time periods. Three groups of intersections lie on this line: those experiencing 7, 3, 2, 1, and 0 collisions in each time period. Such a correlation would mean that the past number of collisions at an intersection in the baseline period would correctly predict the number in the follow-up period. As can be seen, those blue circles are relatively small for intersections with the largest circles being those that represent intersections with 0 or 1 collision. This means that there are relatively few instances of the past number of collisions at an intersection matching with the future number.

The scatter plot shows two deviations from dotted line that indicates where intersections would fall if there were a perfect correlation. In addition to a general scattering in which the number of collisions at an intersection is different for each period, sometimes quite substantially, there is a systematic pattern in the difference between the baseline and follow-up periods. The numbers in the follow-up period trend toward the mean of the follow-up period, which is shown on the dotted "y" line at the value of 1.167, compared to their relation to the mean of the baseline period, shown on the dotted "x" line, positioned at the value of 1.128. This is another illustration of the Regression to the Mean (RTM) discussed earlier. That is one reason for false positives. The "regression line" shown by the dotted red line in the graph illustrates this phenomenon, as its slope is considerably smaller than the perfect correlation line's slope, reflecting this regression downward to the mean.

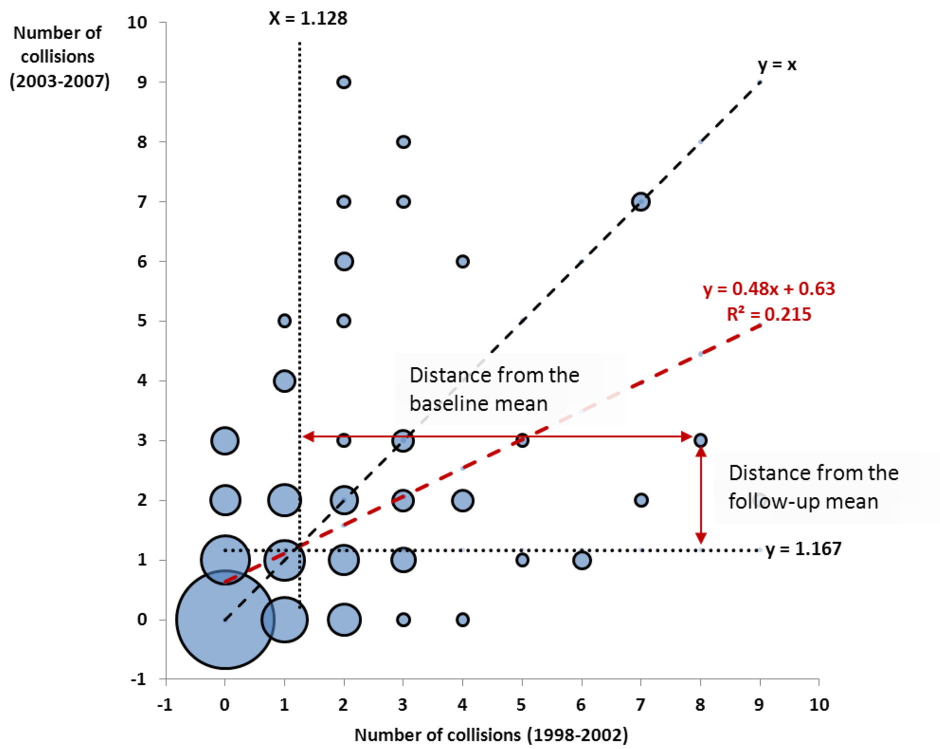


Figure 13. Scatter plot of the number of crashes in the baseline and follow-up period, showing regression toward the mean

Clustering Adjacent Intersections

Having determined that using multiple years creates more accurate benefits than using a single year, we now evaluate the benefits of using the number of crashes in a cluster of adjacent intersections. Another way to think of clustering is to view it as increasing the spatial scale, by combining intersections that are near one another and thereby increasing the number of crashes in the analysis. The assumption is that intersections near one another are likely to share important characteristics such as land-use and traffic volume. For this analysis we grouped 5 adjacent intersections into 36 clusters, and then compared the number of collisions in each cluster for the two time periods (Figure 14).

Figure 14. The 36 clusters of intersections with the number of crashes in the baseline and follow-up period

Performing the same cross tabulation (Figure 15) shows that 3 of the 5 clusters (15 intersections) had between 8 and 9 collisions during the follow-up period.

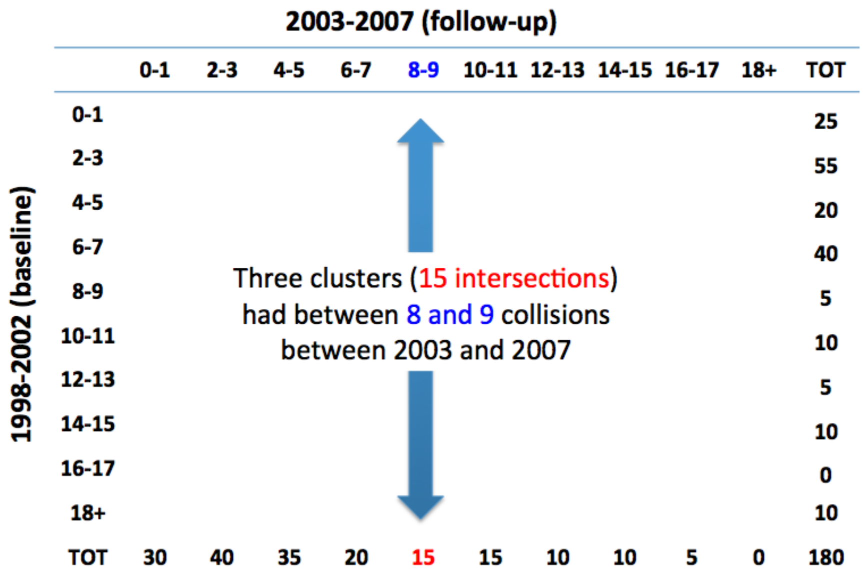


Figure 15. Cross tabulation of the number of crashes in the baseline and follow-up period for clusters of 5 intersections

To determine the rates of true and false positives and negatives, the criterion is set at a level that will result in a number that is comparable to the number of positive intersections, 27, that were identified without clustering. As a result, the criterion for the cluster analysis is 10-11 or more collisions per cluster, which produces 25 intersections, two fewer than we found without clustering.

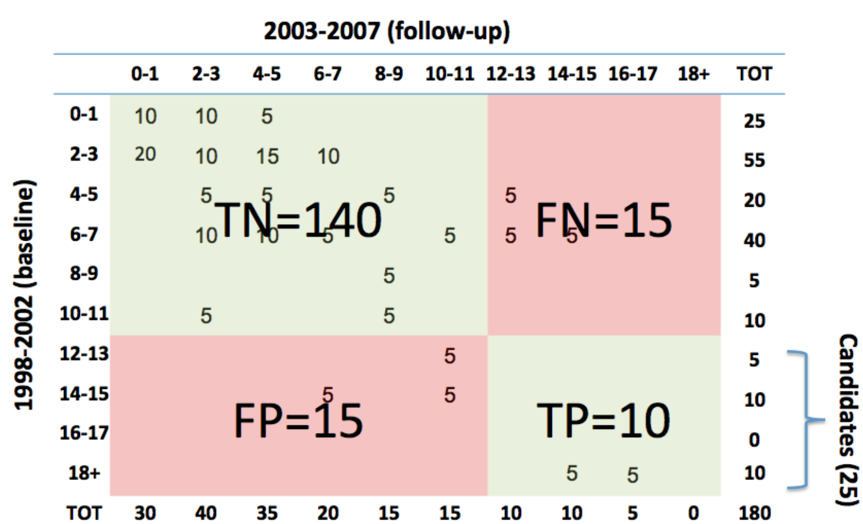


Figure 16. Number of candidate intersections, and the corresponding outcomes, for a threshold of more than 2 collisions per intersection, for clusters of 5 intersections

Figure 17 shows that the 25 candidate intersections were chosen based on 79 collisions. However, in the follow-up period there were only 58 collisions, of which 27 were FP. Of the 124 collision that occurred in non-candidate intersections, about one third were FN (40). That means that about half of the intersection would perform better than the threshold irrespective of any costly inspection and intervention. On the other hand about one third of the intersections that were considered low collision frequency actually did exhibit collision rates that were higher than the threshold. In total for the cluster of 5 intersections the total FP and FN was 67 while for the un-clustered it was 81. This demonstrates the advantage of the clustering of adjacent intersection over studying individual intersections.

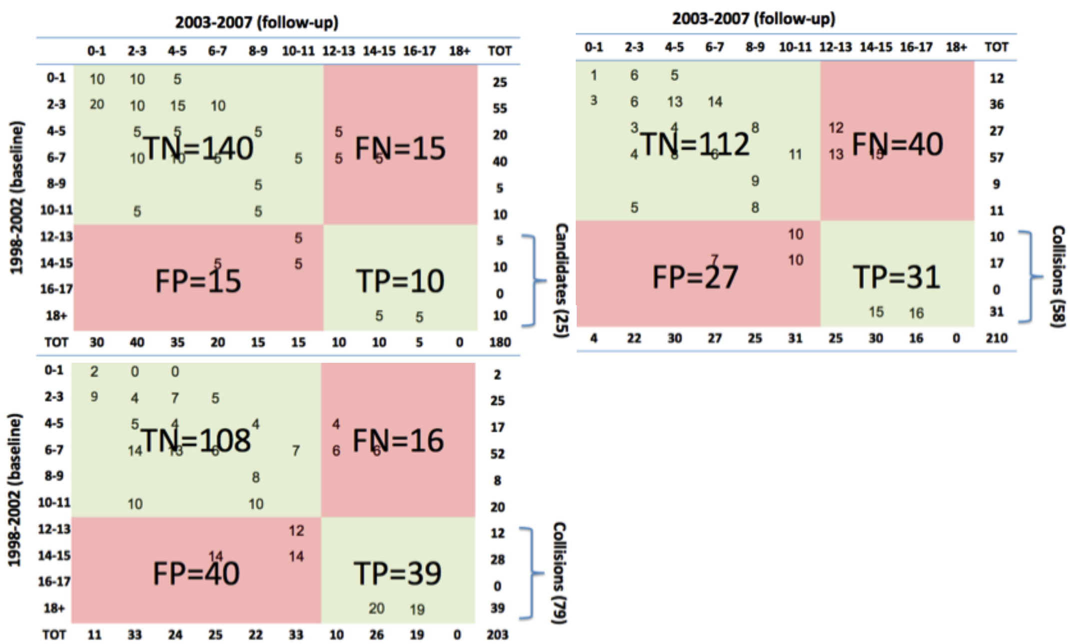


Figure 17. Number of candidate intersections, and the corresponding outcomes, for intersections, collision in the baseline period (bottom-left) and collisions in the follow-up period (top-right), for those with more than 2 collisions at the intersection level for a 5 year sampling period, for clusters of 5 intersections

The *CV(RMSE)* for the clustered has now dropped to 0.69. This is the first time that *CV(RMSE)* dropped below 1, and also this indicates a large improvement in the fit of the data compared to the un-clustered sets (*CV(RMSE)*=1.55).

Clustering Intersections by Common Features

Another approach to improving estimates of the expected number of collision is to combine other information about characteristics of a site with past history. The "Bayesian" framework does exactly that. Under this framework information about the features of individual sites in the network is developed, and used to estimate a "prior probability" for each intersection, that probability is then combined with the actual number of collisions. Similarly it is also possible to group the intersections according to the common features and evaluate whether the estimates of the expected number of collisions is better within each group.

Clustering by Traffic Signals

We begin by using the existence of a traffic signal as a clustering variable. Figure 18 demonstrates the results of such clustering in two scatter plots. The left chart is only for signalized data (52 intersections) and we can see that the average annual collision rate is 2.48, while the right chart is only for non-signalized intersections, which typically have a lower absolute number of crashes, and the average annual collision rate is 0.63.

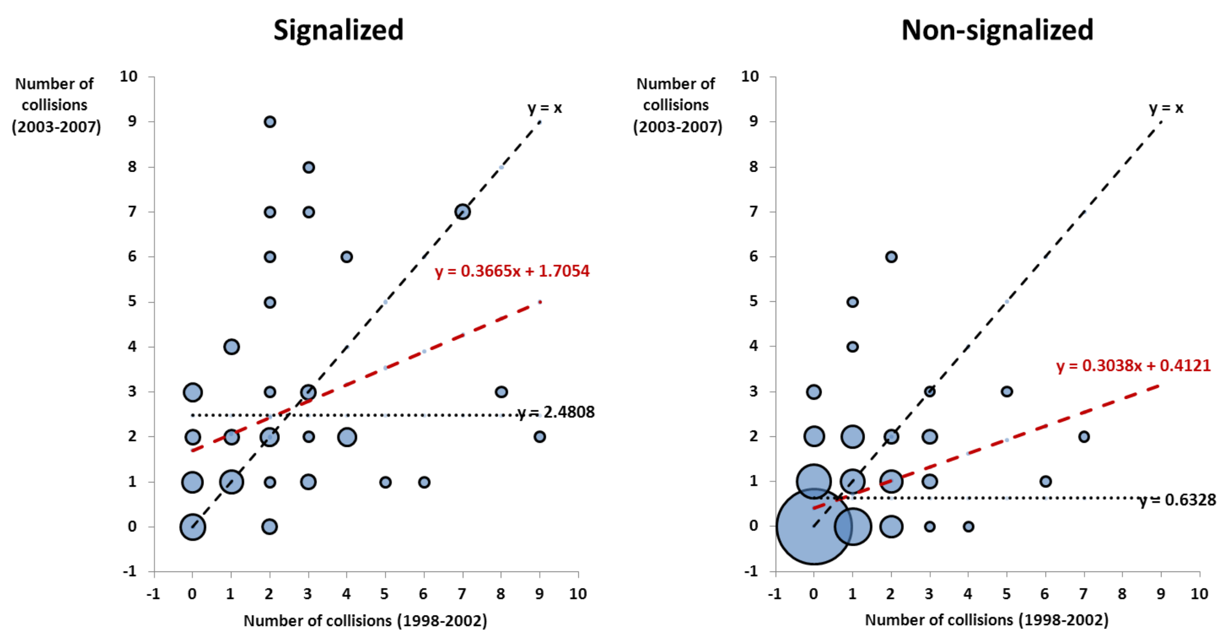


Figure 18. Clustering by traffic signal for a 5 year sampling period

We can see in Figure 18 that the clustering indeed improved the fit on the charts. Moreover, the *CV(RMSE)* has reduced compared to the un-clustered data to 0.96 for the non-signalized and to 1.2 for the signalized intersections to an overall weighted *CV(RMSE)* of 1.03.

Clustering by Multiple Features

It is also possible to cluster by several different features. The list of features and the justification for them is found below.

- Number of baseline collisions

  This variable has been used all along and represents the capability of a previous period to help predict future events.

- Physical proximity

  The distance between two intersections. Intersections that are close to each other are more likely to be clustered together.

- Signalized

  Whether the intersection is signalized.

- Pedestrian attraction

  The level of pedestrian activity around the intersection. Estimated by the number of intersection crossings modeled according to the Alameda County Pedestrian Intersection Crossing Volume Model.[18] Figure 19 shows four alternative models for the number of pedestrian crossings. For the purpose of this clustering, Model 1 was used.

[18] Schneider R.J., L.S. Arnold, and D.R. Ragland. Pilot Model for Estimating Pedestrian: Intersection Crossing Volumes. In Transportation Research Record: Journal of the Transportation Research Board, No. 2140, Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 13– 24.

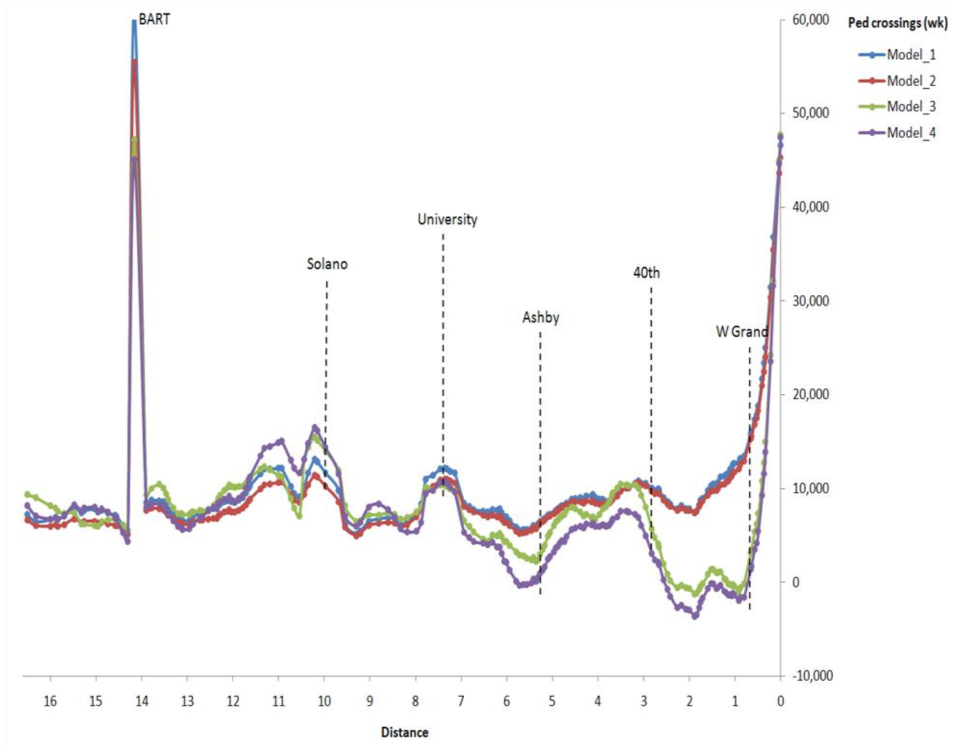


Figure 19. Pedestrian Attractiveness Modeled by 4 Alternative Models of pedestrian intersection crossings per week

Using a two-step clustering algorithm the data was broken to 15 different clusters. The full list of clusters is available in Appendix 4. Examples for 4 of the clusters are shown in Figure 20 and Table 7. The results show that some clusters (e.g., cluster 12) may have a high rate of collisions even though they enjoy a feature which is associated with low crash rates, on the other hand cluster 4 consists of signalized intersections and exhibits a moderate crash rate. These results demonstrate that using a single feature to cluster the data is limited and suggest that further exploring of this approach would be beneficial.

Table 7. Distribution of the number of crashes per intersection for the baseline and follow-up periods

Cluster number	Intersection control	Collisions	Pedestrian
Cluster 12	Non-signalized	Low (0.34)	Moderate (7,479)
Cluster 7	Non-signalized	Moderate (0.63)	High (14,984)
Cluster 4	Signalized	Moderate (0.53)	Moderate-high (9,145)
Cluster 14	Non-signalized	High (2.44)	Moderate (7,286)

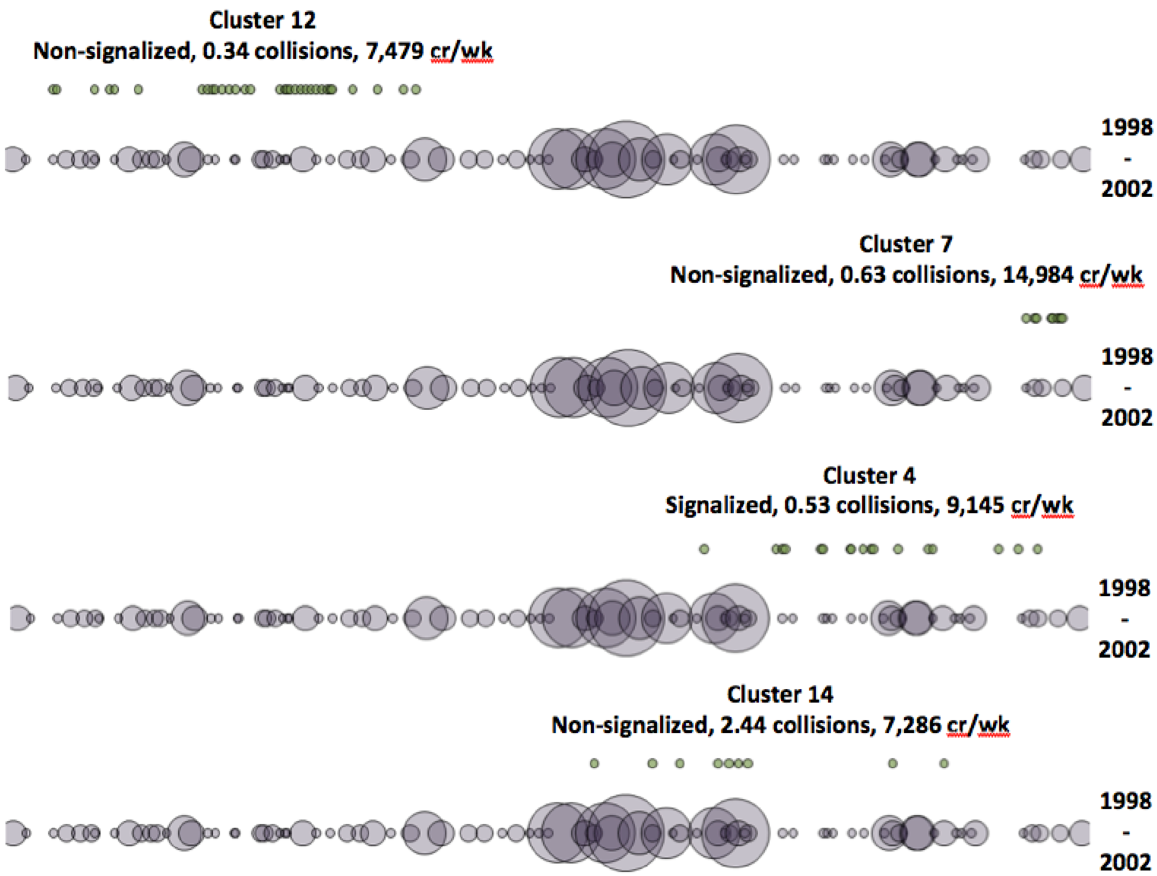


Figure 20. Examples of the Results from Clustering by Multiple Features

Conclusions and Recommendations

This chapter summarizes the results of the report and describes the next steps for the project and for Caltrans.

Conclusions

- Basing decisions on individual intersections and single years is of limited efficacy and will yield substantial numbers of two types of errors: (i) false positives, i.e., selecting a particular intersection for treatment when in fact there would be a fairly small number of collisions there in subsequent years and (ii) false negatives, i.e., failing to select an intersection when in fact that intersection would have a significant number of collisions in ensuing years. If, as is the case, the goal is to prevent the maximum number of collisions per dollar spent, prioritizing spending according to annual crashes at individual intersections would yield limited success.

- Confidence intervals around crash frequencies decrease most rapidly over the first three years, and then decrease more slowly. This may indicate that using a sampling period of three years, as is common, provides a good balance between reducing statistical variation and accounting for changes in the intersections over time. However, for low-crash intersections a longer period, say five years, may be necessary for achieving more stability in estimating the expected number of crashes.

- There are high concentrations of collisions not only for individual intersections, but high concentrations within spatial clusters of intersections, suggesting there are factors affecting not just the individual intersection, but affecting entire sets of intersections. There is an advantage in terms of increased stability of estimated future collisions of clustering adjacent intersections over focusing on individual intersections.

- Table 8 summarizes how well two the baseline and outcome periods conform for the different approaches evaluated using the *CV(RMSE)* measure. The results demonstrate that increasing the compared intervals to 5 years and grouping together 5 adjusted intersections provides a significantly better approximation for predicating pedestrian and bicyclist collisions.

Table 8. *CV(RMSE)* for the different approaches evlauated

Intervals compared	Individual intersections	5 adjacent intersections	Signalized vs. un-signalized
1 YR	3.19		
5 YR	1.55	0.69	1.05

Recommendations

Recommendations for traffic engineers and planners

1. Before starting a safety evaluation, develop a comprehensive database of the area to be considered. The web-based query system of statewide collision data for California developed by SafeTREC (http://www.tims.berkeley.edu/) can be used to help build a collision database. Currently, there is no comprehensive statewide data structure for highway inventory, requiring the information to be developed at the local level.

2. Conduct a descriptive analysis of collision numbers and clusters, followed by an evaluation of sites at different scales with the intent of increasing the stability of estimates of expected injuries through methods described above.

3. Conduct an evaluation of the benefit-cost of different strategies. The tool developed by SafeTREC for the HSIP program is available for this purpose.

Recommendations for further research

In the recent past, data systems and methods have been developed for conducting analyses and implementing strategies for improving pedestrian and bicyclist safety. However, work is needed to refine and extend these data systems and methods. Some of the needed research includes the following:

1. Develop methods of analyses based on Bayesian techniques to estimate expected injuries. This will require constructing databases with infrastructure inventory data and reliable estimates of vehicle and pedestrian volume.

2. Identify specific features of the infrastructure associated with pedestrian and bicyclist risk, thereby permitting strategies that focus on sites with these specific features and increase the benefits of Bayesian analysis.

3. Identify techniques which will allow to cluster intersections by multiple features.

4. Explore the implications and benefit-cost of strategies above the specific site-level, specifically, those involving extended street segments, clusters of intersections, or systemic approaches.

5. Develop tools for evaluating the impact of pedestrian and bicyclist injury countermeasures on Level of Service (LOS) for vehicle traffic. This is important for being able to provide optimal level of service for all modes of traffic.

IMPLEMENTATION STRATEGIES

Strategies for implementation include: (i) provide training for traffic engineers and planners, (ii) design Internet-based or stand-alone tools that incorporate methods developed in this study as well other available methods; (iii) conduct applied research to extend and refine the methods and approaches developed in this project, and, finally, (iv) propose that methods and approaches for identifying sites with optimal potential be incorporated as one of the goals of the Strategic Highway Safety Plan (SHSP) process.

Appendices

Appendix 1. Photographs of VRU arterials along San Pablo Avenue

(a) (b) (c) (d) (e) (f)

Figure 1A. VRU arterials crossing San Pablo Ave. along the Study Area
(a) Hill St. (b) Solano Ave. (c) University Ave. (d) Ashby Ave. (e) 40th/Adeline (f) 17th St.

Appendix 2. VRU crashes for intersections along San Pablo Ave. from 1998 to 2007

INTERSECTION	98	99	00	01	02	03	04	05	06	07	TOTAL
SOLANO AVE (N)	1	0	1	1	0	0	0	1	0	0	**4**
CLINTON AVE	0	0	0	0	1	0	0	2	0	0	**3**
RT 80 (N)	0	0	0	0	0	0	0	0	0	0	**0**
ROOSEVELT AVE	0	1	0	0	0	0	0	0	0	0	**1**
RT 80	0	0	0	0	0	0	0	0	0	0	**0**
BARRETT AVE	0	0	0	0	2	0	0	0	0	0	**2**
NEVIN AVE	0	1	0	1	0	0	0	1	0	0	**3**
MACDONALD AVE	0	0	1	1	0	0	1	0	0	0	**3**
BISSELL AVE	1	0	0	0	0	0	0	0	0	0	**1**
OHIO ST	0	0	0	0	0	0	0	0	0	0	**0**
CONLON AVE	0	0	0	1	0	0	0	0	0	0	**1**
ROSE	0	0	0	0	0	0	1	0	0	0	**1**
WALL AVE	0	1	2	0	0	0	0	1	1	0	**5**
KNOTT AVE (N)	0	0	0	0	0	0	0	0	0	0	**0**
KNOTT AVE	0	0	1	0	1	0	0	0	1	0	**3**
CUTTING BLVD	0	0	1	1	0	1	2	1	2	0	**8**
HUBER AVE	0	0	0	2	0	0	0	0	0	0	**2**
HILL ST	0	0	0	0	1	0	0	1	0	0	**2**
BLAKE ST (N)	3	0	0	1	0	0	0	0	0	0	**4**
POTRERO AVE	2	0	0	0	1	0	0	1	0	2	**6**
CYPRESS AVE	0	0	0	0	0	0	0	0	0	0	**0**
CARLOS AVE	1	0	0	0	0	0	0	0	0	0	**1**
MADISON AVE (N)	0	0	0	0	0	1	0	1	0	1	**3**
MADISON AVE	0	0	0	1	0	0	0	0	0	0	**1**
JEFFERSON AVE	0	0	0	0	0	0	0	0	0	1	**1**
ALAMEDA AVE	0	0	0	0	0	0	0	1	0	0	**1**
BAYVIEW AVE	0	0	0	1	0	1	1	0	1	1	**5**
WENK AVE	0	0	0	1	0	1	0	0	0	1	**3**
ORCHARD AVE	0	0	0	0	0	0	0	0	0	0	**0**
SCHMIDT LN	0	0	0	0	0	0	0	0	0	0	**0**
TEHAMA AVE	0	0	0	0	0	1	0	0	0	0	**1**
PORTOLA DR	0	1	1	0	0	0	0	0	1	0	**3**
BURLINGAME AVE	0	0	2	0	0	0	0	0	0	0	**2**
MOESER LN	0	1	1	0	0	1	0	1	0	0	**4**
PLUMAS AVE	0	1	0	0	0	0	0	0	0	0	**1**
SUTTER AVE	0	0	0	0	1	0	0	0	0	0	**1**
WALDO AVE	0	0	0	1	0	0	0	0	0	0	**1**

SANTA CRUZ AVE	0	0	0	0	0	0	0	0	0	0	**0**
HUNTINGTON AVE	0	0	0	0	0	0	0	0	0	0	**0**
PANAMA AVE	0	0	0	0	0	0	0	0	0	0	**0**
STOCKTON AVE	0	0	0	0	3	0	1	0	0	0	**4**
SACRAMENTO AVE	0	0	0	0	0	0	0	0	0	0	**0**
FRESNO AVE	0	0	0	0	0	0	0	0	0	0	**0**
COLUMBIA AVE	0	1	0	0	0	0	0	1	1	0	**3**
VAN FLEET AVE	0	0	0	0	0	0	0	0	1	0	**1**
SAN JOSE AVE	0	0	0	0	0	0	0	0	0	0	**0**
LINCOLN AVE	0	0	0	0	1	0	1	1	1	2	**6**
EL DORADO AVE	0	0	0	0	0	0	0	0	0	0	**0**
CENTRAL AVE	0	0	0	1	1	1	0	1	0	0	**4**
SAN DIEGO ST	0	0	0	0	0	0	0	0	0	0	**0**
FAIRMOUNT AVE	1	1	0	0	0	3	1	2	0	1	**9**
CARLSON BLVD	1	0	2	0	0	1	0	1	0	1	**6**
KAINS AVE	0	0	0	0	0	0	0	0	0	0	**0**
BRIGHTON AVE	0	0	0	0	1	0	1	1	2	0	**5**
CLAY ST (N)	0	0	0	0	0	0	0	0	0	1	**1**
GARFIELD AVE	0	0	0	0	0	0	0	0	0	0	**0**
CASTRO ST (N)	0	0	0	2	0	0	0	0	1	0	**3**
PORTLAND AVE	0	0	0	0	0	0	0	0	0	0	**0**
WASHINGTON AVE (N)	1	2	1	1	0	0	0	0	0	1	**6**
WASHINGTON AVE	0	0	0	0	0	0	0	0	0	0	**0**
SOLANO AVE	1	2	0	0	0	1	1	2	3	1	**11**
BUCHANAN ST	0	0	0	0	0	0	0	1	0	0	**1**
MARIN AVE	0	1	0	1	0	1	0	1	0	0	**4**
MONROE ST	0	0	0	2	0	0	0	0	0	0	**2**
DARTMOUTH ST	0	0	0	0	0	0	0	0	0	0	**0**
HARRISON ST	0	0	1	0	0	0	1	0	0	3	**5**
GILMAN ST	1	0	1	0	0	1	0	3	1	0	**7**
CAMELIA ST	1	0	0	0	0	1	0	0	0	0	**2**
PAGE ST	1	0	0	0	0	0	0	0	1	1	**3**
JONES ST	1	0	0	0	0	0	0	0	0	1	**2**
CEDAR ST	1	1	4	0	1	0	0	1	2	4	**14**
VIRGINA ST	1	3	2	1	0	1	0	1	0	0	**9**
& FRANCISCO ST	0	0	0	0	0	0	0	0	0	0	**0**
DELAWARE ST	2	1	0	0	0	2	0	4	0	1	**10**
HEARST AVE	1	1	0	0	0	0	0	0	0	0	**2**
UNIVERSITY AVE	0	3	0	2	2	1	0	2	2	2	**14**
ADDISON ST (N)	2	0	1	0	1	0	2	3	0	1	**10**
ADDISON ST	0	0	0	0	0	0	0	0	0	0	**0**

COWPER ST	0	0	0	0	0	0	0	0	0	0	**0**
ALISTON WAY	3	1	1	3	1	0	1	0	0	1	**11**
BANCROFT WAY	1	1	2	0	1	0	1	0	2	0	**8**
CHAUCER ST	0	0	0	0	0	0	0	0	0	0	**0**
CHANNING WAY	0	0	2	0	0	1	1	1	1	2	**8**
DWIGHT WAY	0	1	1	2	2	0	0	1	0	0	**7**
BLAKE ST	0	0	0	0	1	0	0	0	0	0	**1**
PARKER ST	0	0	1	1	0	2	0	0	0	0	**4**
CARLETON ST (N)	0	0	0	0	0	0	0	0	0	0	**0**
CARLETON ST	0	0	0	0	0	2	0	0	0	0	**2**
DERBY ST	0	0	0	0	0	0	0	0	0	0	**0**
PARDEE ST	0	0	1	0	0	0	0	0	0	0	**1**
WARD ST	0	0	0	0	0	0	0	0	0	0	**0**
GRAYSON ST	0	0	0	0	0	0	0	0	1	0	**1**
OREGON ST	1	1	1	3	0	0	1	0	0	0	**7**
HEINZ AVE	1	1	0	1	0	0	0	0	0	0	**3**
RUSSELL ST	0	1	0	0	0	0	0	0	1	0	**2**
BURNETT ST	2	0	0	0	0	0	0	0	0	0	**2**
ASHBY AVE	1	1	1	1	4	0	0	2	0	1	**11**
MURRAY ST	0	0	2	0	1	0	0	0	2	0	**5**
CARRISON ST	0	0	0	0	0	0	0	0	0	0	**0**
FOLGER AVE	0	0	0	0	1	0	0	1	0	1	**3**
HASKELL ST	1	0	1	0	0	0	0	0	0	1	**3**
67TH ST (N)	0	0	0	0	0	1	0	0	0	1	**2**
67TH ST	0	0	0	0	0	0	0	0	0	0	**0**
66TH ST (N)	0	0	0	0	0	0	1	0	0	1	**2**
66TH ST	0	0	0	0	0	0	0	0	0	0	**0**
65TH ST	0	0	0	0	0	0	0	0	0	0	**0**
PEABODY LN	0	0	0	0	0	0	1	0	0	0	**1**
OCEAN AVE	0	0	0	0	0	0	0	0	0	0	**0**
ALCATRAZ AVE	0	0	0	0	0	0	1	0	0	1	**2**
64TH ST	0	0	0	0	0	0	0	0	0	0	**0**
63RD ST (N)	0	0	0	1	0	0	0	1	0	0	**2**
63RD ST	0	0	0	0	0	0	0	0	0	0	**0**
62ND ST	0	0	0	0	1	0	0	0	0	1	**2**
61ST ST	0	0	0	0	0	1	0	0	0	0	**1**
60TH ST	0	0	0	0	0	0	0	0	0	0	**0**
59TH ST (N)	0	0	0	0	0	0	0	0	0	0	**0**
59TH ST	0	0	0	0	0	0	0	0	0	0	**0**
STANFORD AVE (N)	0	0	0	0	0	0	0	0	0	0	**0**
STANFORD AVE	0	0	1	0	0	0	0	0	1	0	**2**

57TH ST	0	0	1	0	0	0	0	0	0	0	**1**
AILEEN ST	1	0	0	0	0	0	0	0	0	0	**1**
56TH ST	0	0	0	0	0	0	0	1	0	0	**1**
55TH ST (N)	0	0	0	0	0	0	0	0	0	0	**0**
55TH ST	0	0	0	0	0	0	0	0	0	0	**0**
54TH ST (N)	0	0	0	0	0	0	0	0	0	0	**0**
54TH ST	0	0	0	0	0	0	0	0	0	0	**0**
53RD ST (N)	0	0	0	0	0	0	0	0	0	0	**0**
53RD ST	1	0	0	0	0	0	0	0	1	0	**2**
48TH ST	0	0	0	0	0	0	0	0	0	0	**0**
47TH ST	0	0	1	0	0	0	1	1	0	0	**3**
45TH ST (N)	0	0	0	0	0	0	0	0	1	1	**2**
45TH ST	0	0	0	0	0	0	0	1	1	1	**3**
43RD ST	0	0	0	1	0	0	0	1	0	0	**2**
PARK AVE	0	0	1	2	1	0	0	1	0	1	**6**
41ST ST	0	0	0	2	1	0	0	1	0	0	**4**
40TH ST	0	0	0	0	2	1	0	0	2	0	**5**
YERBA BUENA AVE	0	0	0	0	0	0	0	0	0	0	**0**
PARKING LOT	0	0	0	0	0	0	0	1	0	0	**1**
ADELINE ST	1	0	2	0	1	0	0	0	1	1	**6**
W MACARTHUR BLVD	1	0	2	0	1	0	0	0	1	1	**6**
37TH ST	0	0	0	0	0	0	0	0	0	0	**0**
36TH ST	0	0	0	0	0	0	0	1	0	2	**3**
35TH ST	0	0	1	0	0	1	0	0	0	0	**2**
34TH ST	1	0	2	0	0	0	1	0	0	2	**6**
33RD ST	0	0	0	0	0	0	1	0	0	0	**1**
FILBERT ST	0	0	1	0	0	0	1	0	0	0	**2**
32ND ST	0	0	1	0	0	1	0	0	1	0	**3**
MYRTLE ST	0	0	0	0	0	0	0	0	0	1	**1**
31ST ST	0	0	0	0	1	0	0	0	0	0	**1**
MARKET ST (N)	1	0	1	1	0	1	0	0	0	1	**5**
30TH ST	0	0	0	0	0	2	1	0	0	0	**3**
MARKET ST	0	0	0	0	0	0	0	0	0	0	**0**
29TH ST	0	0	0	0	0	0	0	0	0	0	**0**
28TH ST (N)	0	0	0	0	0	0	0	0	0	0	**0**
28TH ST	0	0	0	0	0	0	0	0	0	0	**0**
27TH ST (N)	0	0	0	0	0	1	0	0	0	2	**3**
27TH ST	0	0	0	0	0	0	0	0	0	0	**0**
MILTON ST	0	0	0	0	0	0	0	0	0	0	**0**
MEAD AVE	0	0	0	0	0	0	0	0	0	0	**0**
SYCAMORE ST	0	0	0	0	0	0	1	1	0	0	**2**

ATHENS AVE	0	0	0	0	0	0	0	0	0	0	**0**
25TH ST	0	0	0	0	0	0	0	0	0	0	**0**
WEST ST	0	0	0	0	0	0	0	0	0	0	**0**
ISABELLA ST	0	0	0	0	1	0	0	0	0	0	**1**
24TH ST	0	0	0	0	0	0	0	0	0	0	**0**
23RD ST	1	0	0	0	1	2	0	0	0	0	**4**
BRUSH ST	0	0	0	0	0	1	0	0	0	0	**1**
W GRAND AVE	0	1	0	0	1	4	0	2	0	3	**11**
CASTRO ST	0	0	0	0	0	0	0	0	0	0	**0**
CASTRO ST (S)	0	0	0	0	0	0	0	0	0	0	**0**
21ST ST	0	0	0	0	0	0	0	0	0	0	**0**
MARTIN LUTHER KING JR	0	0	0	0	0	0	0	0	0	0	**0**
20TH ST	0	0	0	2	0	0	0	0	0	0	**2**
WILLIAM ST	0	0	0	0	0	0	1	0	0	0	**1**
19TH ST (N)	0	0	0	0	0	0	0	0	0	0	**0**
19TH ST	0	0	0	0	0	0	0	0	0	0	**0**
18TH ST	0	0	0	0	0	0	0	0	0	0	**0**
CLAY ST	0	0	2	0	1	0	0	0	1	0	**4**
16TH ST	0	0	0	0	0	0	0	0	0	0	**0**
FRANK H OGAWA PLAZA	0	0	0	0	0	0	0	0	0	0	**0**
TOTAL	**40**	**30**	**51**	**42**	**40**	**40**	**27**	**53**	**38**	**52**	**413**

Appendix 3. Literature Review: Methodologies for Identifying High Priority Pedestrian Locations

Resource Allocation for Pedestrian Safety Programs: Methodologies for Identifying High Priority Pedestrian Locations

PREPARED BY

Tierra Bills, David R. Ragland and Staff
University of California
SafeTREC, Safe Transportation Research & Education Center

May, 2011

Methodologies for Identifying High Priority Pedestrian Locations

Objectives

The objectives of this project are to:

- Review methodologies for identifying high priority pedestrian locations
- Provide insight for agencies that are developing pedestrian safety programs

Resource Allocation

Agencies must determine how to allocate limited resources to achieve the greatest impact from traffic safety programs. SafetyAnalyst[19] was developed as a cooperative effort by FHWA and participating state and local agencies and provides analytical tools for use in the decision-making process to identify and manage a system-wide program of site-specific improvements to enhance highway safety by cost-effective means. Identifying high priority locations involves the following:

- Network Screening: to identify locations with high crash frequency and high potential for crash reduction
- Diagnosis
- Countermeasure Selection
- Economic Appraisal
- Priority Ranking
- Countermeasure Evaluation

Location Categories

This brief summary focuses on network screening. Typically, identifying locations high potential for crash reduction has centered on “spot” locations involving individual intersection or discrete sites. However, here we compare these different scales: spot locations, corridors and zones.

- **Sites**
  Spot locations including individual intersections and non-intersections
- **Corridors**
  Roadway sections of 0.5 to 5 miles in length
- **Zones**
  Target areas that can range from a single neighborhood or business to an entire jurisdiction

[19] SafetyAnalyst. *SafetyAnalyst Overview*. http://www.safetyanalyst.org/ [accessed June 2, 2011].

Sites

Sites or spot locations are defined as individual intersections and non-intersections. This includes ramps, road segments, or any distinct location where crash may take place. An example of site identification is a study conducted by Morency and Cloutier[20] to illustrate the geographic distribution of pedestrian crash sites in the urban setting of Montreal, Canada using an alternative data source.

Sites: Identification Example

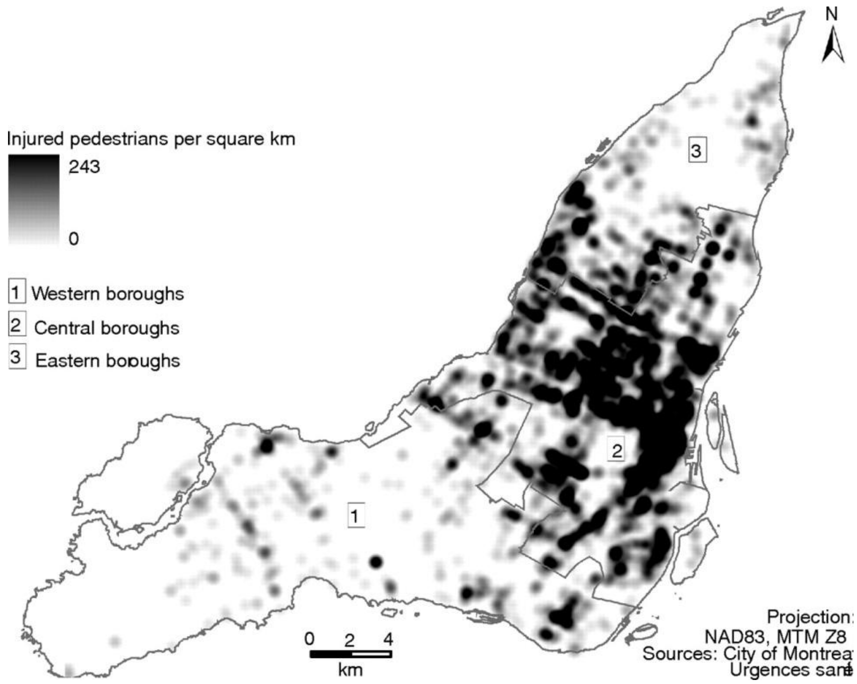
- Data on pedestrian victims in Montreal were extracted for a 5-year period (1999-2003) from ambulance services information systems.
- The locations of crash sites and pedestrian victim density were mapped using a geographic information system.
- Pedestrian "black spots" were defined as sites where there had been at least 8 pedestrian victims.
- The results identified 22 "black spots" representing only 1 percent of city intersections with at least one victim and 4 percent of all injured pedestrians. The number and population rates of injured pedestrians are greater in central boroughs. Over the 5-year period, in some central boroughs of Montreal, pedestrian crashes occurred at up to 26 percent of intersections.
- The study concluded that most pedestrians were injured at locations that would have been missed by the black spot approach and that prevention strategies should include comprehensive environmental measures such as global reduction of traffic volume and speed.

[20] Morency, P et al. From targeted "black spots" to area-wide pedestrian safety. *Inj Prev* 2006;12:360-364.

Distribution of Pedestrian Victims in Rosemont/Petite-Patrie in Montreal According to Urgences-Sante (1999-2003)

Source: Morency, P et al. From targeted "black spots" to area-wide pedestrian safety. *Inj Prev* 2006;12:360-364.

Density Mapping of Pedestrian Victims



Source: Morency, P et al. From targeted "black spots" to area-wide pedestrian safety. *Inj Prev* 2006;12:360-364.

SITES: Strengths and Weaknesses

STRENGTHS	WEAKNESSES
• Intuitive, methods exist to identify sites • Attacking individual problem sites	• Fails to address the area-wide problem • Efficiency is limited when looking at high density scattered crashes • Instability of estimates of expected injuries, especially if injury

Corridors

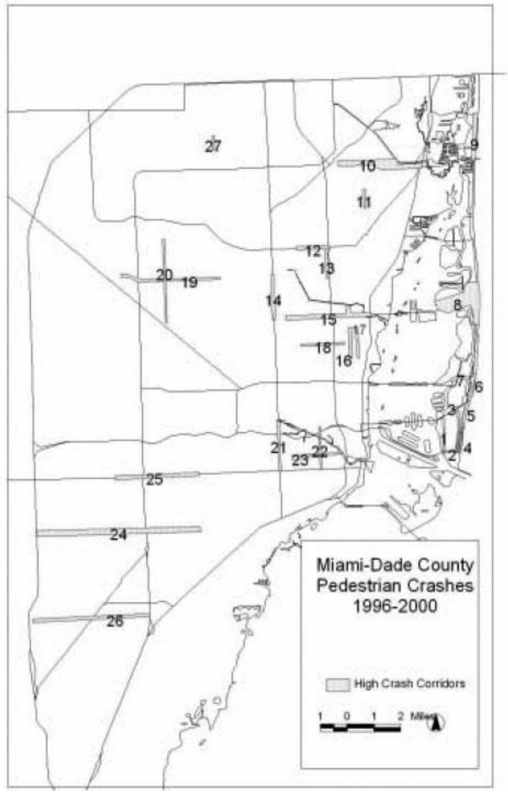
Corridors are defined as roadway sections of 0.5 miles to 5 miles. Corridors may also include parallel segments on either or both sides of the roadway. An example of corridor identification is a study conducted by the U.S. Department of Transportation [21] to identify high crash corridors in Miami Dade County, Florida.

Corridors: Identification Example

- All pedestrian crashes between 1996-2000 were extracted from the FDOT records and mapped
- Crash density was evaluated
- The analysis identified 27 high crash corridors
- Each corridor was then ranked by crash index and corridors with a crash index of 6 or more were selected for treatment
- The twelve selected corridors accounted for 12 percent of all crashes and 14 percent of fatal crashed in Miami Dade County
- Selected corridors were analyzed using the Pedestrian and Bicycle Crash Analysis Tool (PBCAT) and the data were merged back into the GIS spreadsheet with demographic information for each crash
- The collected data was compiled into a guide listing crashes by location in each corridor
- Countermeasures were selected for each corridor
- Outreach and awareness campaigns to address particular safety problems will be implemented in to future

[21] U.S. Department of Transportation. Federal Highway Administration. 2002. *Pedestrian Safety Engineering and Intelligent Transportation System-Based Countermeasures Program for Reduced Pedestrian Fatalities, Injuries, Conflicts and Other Surrogate Measures, Final Problem Identification, Countermeasure Selection, and Outreach & Awareness Report.* University of Florida Transportation Research Center and Department of Urban and Regional Planning.

Miami Dade High Crash Corridors



Source: U.S. Department of Transportation. Federal Highway Administration. 2002. *Pedestrian Safety Engineering and Intelligent Transportation System-Based Countermeasures Program for Reduced Pedestrian Fatalities, Injuries, Conflicts and Other Surrogate Measures, Final Problem Identification, Countermeasure Selection, and Outreach & Awareness Report.* University of Florida Transportation Research Center and Department of Urban and Regional Planning.

CORRIDORS: Strengths and Weaknesses

STRENGTHS	WEAKNESSES
• Considered different pedestrian crash types • Gain numbers and therefore increase stability of estimates of	• Much detail is lost • Little has been done to test out the analytical capabilities • Less effective when looking at intersections of like sites • Clusters or scattered crashes • Need to spread countermeasures over a greater area or number of sites.

Zones

A zone is a targeted area of interest that may be small as a single neighborhood, to as large as an entire jurisdiction. The zone process provides a method of targeting pedestrian safety improvements in a cost efficient manner. Examples of zone identification include two studies conducted by Pulugurtha, and Nambisan[22,23] to identify pedestrian high crash zones in the Las Vegas Metropolitan Area.

Zones: Identification Examples

- Identify problems for analyses
- Map pedestrian crash data; zones may be linear or circular in shape
- Overlay pedestrian crash data on the zones' coverage
- Compute pedestrian crash rates and rank the zones
- Identify high crash locations in the selected zones
- Rank high crash zones

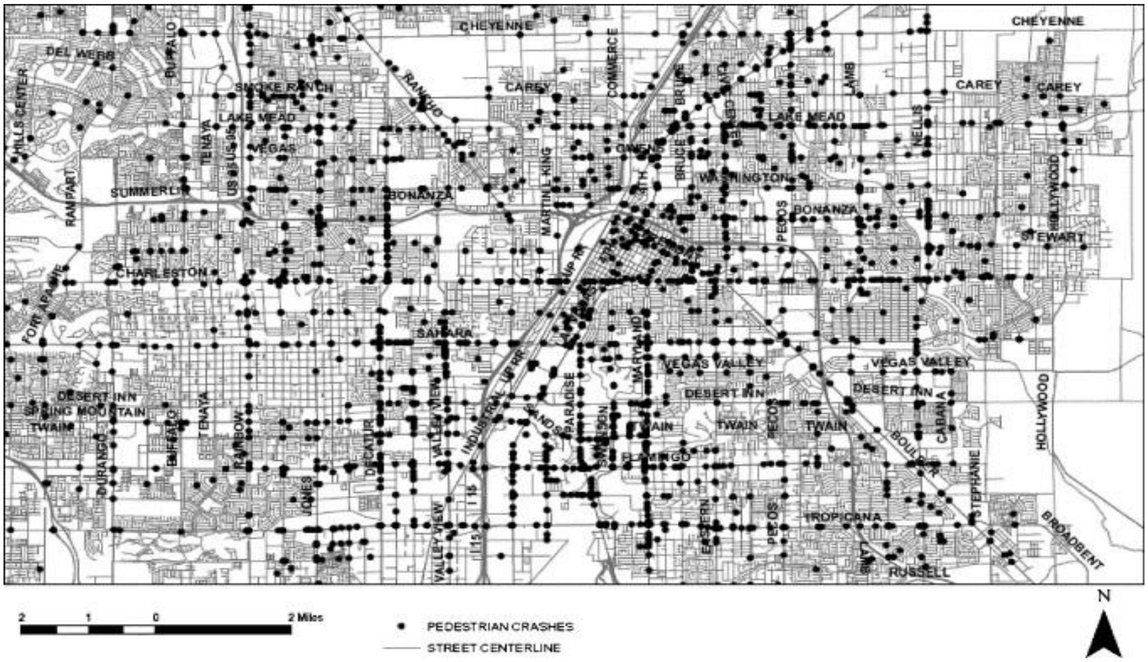
[22] Pulugurtha, S. S. and Nambisan, S. S. (2003). *A Methodology to Identify High Pedestrian Crash Locations: An illustration using the Las Vegas metro area. Las Vegas*

[23] Pulugurtha, S. S. Krishnakumar, V. K. and Nambisan, S.S. (2007). New methods to identify and rank high pedestrian crash zones: An illustration *Accident Analysis & Prevention* 39 (4): 800-811.

- Results obtained from the evaluation of methods to rank high pedestrian crash zones indicated a significant variation in ranking when individual methods were considered

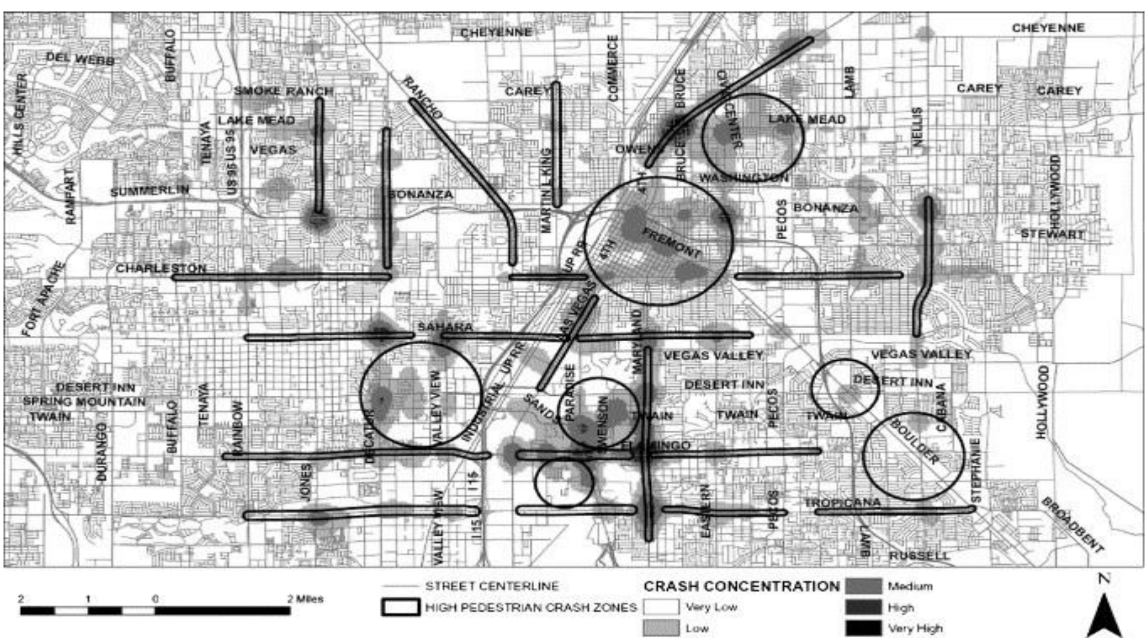
- Recommendations suggest using composite methods in ranking high pedestrian crash zones instead of individual methods

Spatial Distributions of Pedestrian Crashes in the Las Vegas Metropolitan Area (1998–2002)



Source: Pulugurtha, S. S. and Nambisan, S. S. (2003). *A Methodology to Identify High Pedestrian Crash Locations: An illustration using the Las Vegas metro area. Las Vegas*

High pedestrian crash zones in the Las Vegas metropolitan area



Source: Pulugurtha, S. S. and Nambisan, S. S. (2003). *A Methodology to Identify High Pedestrian Crash Locations: An illustration using the Las vegas metro area. Las Vegas*

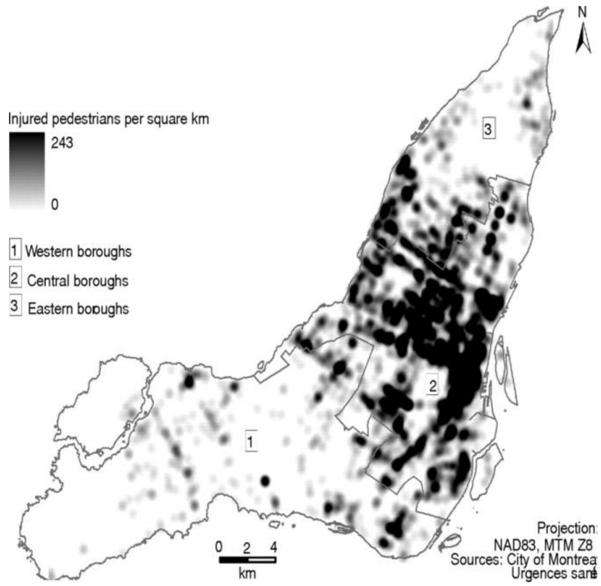
ZONES: Strengths and Weaknesses

STRENGTHS	WEAKNESSES
• Maximizes the size injuries addressed	• Less effective when dealing with individual locations

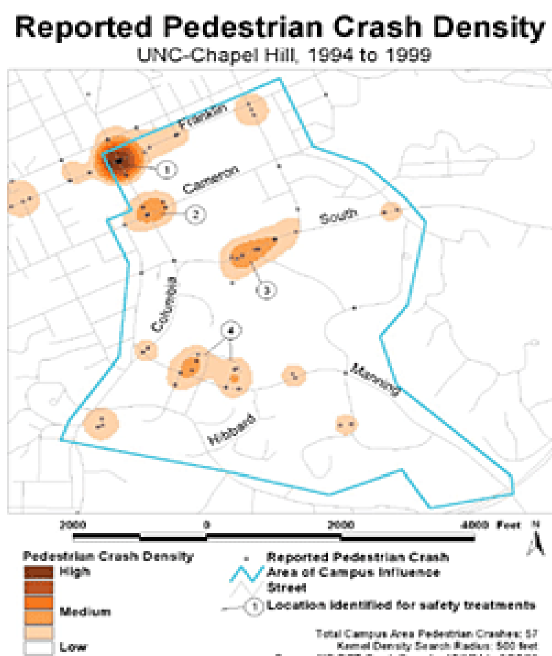
Conclusion

The methodologies vary based on different circumstances and as a result there are a variety of approaches for allocating resources for pedestrian safety programs. Different levels of scale have different strengths and weaknesses and should be chosen based on the purpose of the investigation.

Sample: Density Maps



Morency, P et al. From targeted “black spots” to area-wide pedestrian safety. *Inj Prev* 2006;12:360-364.



Source: Schneider, R. J., R. M. Ryznar, et al. (2004). *An accident waiting to happen: a spatial approach to proactive pedestrian planning.*

References

1. U.S. Department of Transportation. (1998). *Zone Guide for Pedestrian safety*.
2. (2001). *An analysis of North Carolina Guidelines and Criteria for Establishing School Walk Zones North Carolina State University.*
3. (2006). *Federal Highway Administration University Course on Bicycle and Pedestrian Transportation.*
4. Carter, D. L., W. W. Hunter, et al. (2006). *Pedestrian and Bicyclist Intersection Safety: Final report.* McLean, VA, Turner-Fairbank Highway research center: 96.
5. Center, U. o. F. T. R. (2002). *Final Problem Identification, Countermeasure selection, and Outreach & Awareness Report.*
6. Cottrell, W. D. and S. Mu (2005). *Utah Intersection Safety Recurrent Crash Sites: Identification, Issues and Factors.* Salt lake city, University of Utah.
7. Geurts, K., I. Thomas, et al. (2005). "Understanding spatial concentrations of road accidents using frequent item sets." *Accident Analysis & Prevention* 37(4): 787-799.
8. Geurts, K., G. Wets, et al. (2003). "Profiling of High-Frequency Accident Locations by Use of Association Rules." *Transportation Research Record: Journal of the Transportation Research Board* 1840(-1): 123-130.
9. Lyon, C., B. Gotts, et al. (2007). "Comparison of Alternative Methods for Identifying Sites with High Proportion of Specific Accident Types." *Transportation Research Record: Journal of the Transportation Research Board* 2019(-1): 212-218.
10. Lyon, C. and B. Persaud (2002). "Pedestrian Collision Prediction Models for Urban Intersections." *Transportation Research Record: Journal of the Transportation Research Board* 1818(-1): 102-107.
11. Morency, P. and M. S. Cloutier (2006). "From targeted "black spots" to area-wide pedestrian safety." *Inj Prev* 12(6): 360-364.
12. Moudon, A. V. (2001). *Targeting Pedestrian Infrastructure Improvements: A Methodology to Assist Providers in Identifying Suburban Locations with Potential Increases in Pedestrian Travel.* Seattle, WA, University of Washington.
13. Plazak, D. J. and R. R. Souleyrette (2003). *Process to Identify High Priority Corridors for Access Management Near urban Areas in Iowa Using Spatial Data*. Ames, IA, Iowa State University, Center for Transportation Research and Education: 12.
14. Pulugurtha, S. S., V. K. Krishnakumar, et al. (2007). "New methods to identify and rank high pedestrian crash zones: An illustration." *Accident Analysis & Prevention* 39(4): 800-811.
15. Pulugurtha, S. S. and S. S. Nambisan (2003). *A Methodology to Identify High Pedestrian Crash Locations: An illustration using the Las Vegas metro area.* Las Vegas, University of Nevada, Las Vegas.
16. Ragland, D. R. and C.-Y. Chan (2008). *High Collision Concentration Location Table C Evaluation and Recommendations.*

17. Rees, J. (2003). *Corridor Management: Identifying Corridors with Access Problems and Applying Access Management Treatments,* A U.S. 20 Study, Iowa State University: 32.
18. Richard, K. R., D. J. Torbic, et al. (2007). *Overview of SafetyAnalyst: Software Tools for Safety Management of Specific Highway Sites.*
19. Schneider, R., A. Khattak, et al. (2001). "Method of Improving Pedestrian Safety Proactively with Geographic Information Systems: Example from a College Campus." *Transportation Research Record: Journal of the Transportation Research Board* 1773(-1): 97-107.
20. Schneider, R. J., R. M. Ryznar, et al. (2004). "An accident waiting to happen: a spatial approach to proactive pedestrian planning." *Accident Analysis & Prevention* 36(2): 193-211.
21. SEMCOG (1997). SEMCOG Traffic Safety Manual. Detroit
22. Zegeer, C. V. and L. Sandt. H*ow to Develop a Pedestrian Safety Action Plan*, University of North Carolina: 159.
23. Zemotel, L. M. and D. K. Montebello (2000). *Interregional corridors : a guide for plan development and corridor management*. [St. Paul, Minn.], Minnesota Dept. of Transportation.

Appendix 4. Results of the Two-Step clustering algorithm

Clusters

Feature Importance
1.0 0.8 0.6 0.4 0.2

Cluster	10	12	11	4	1	14	3	7	13	2	5	8	9	15	6
Label															
Description															
Size	19.4% (35)	17.8% (32)	14.4% (26)	9.4% (17)	7.8% (14)	5.0% (9)	4.4% (8)	4.4% (8)	4.4% (8)	3.9% (7)	2.8% (5)	2.2% (4)	1.7% (3)	1.7% (3)	0.6% (1)
Features	Model_2 9.048.8650	Model_2 7.479.9309	Model_2 7.142.2348	Model_2 9.145.3455	Model_2 7.532.0074	Model_2 7.286.9615	Model_2 7.240.2254	Model_2 14.984.2588	Model_2 6.815.3629	Model_2 9.975.9479	Model_2 8.068.4926	Model_2 24.447.7188	Model_2 41.468.7487	Model_2 6.860.9890	Model_2 55.436.0178
	Distance 2.72	Distance 12.83	Distance 6.60	Distance 3.37	Distance 12.25	Distance 5.33	Distance 13.04	Distance 0.72	Distance 13.60	Distance 5.08	Distance 6.95	Distance 0.33	Distance 0.06	Distance 6.90	Distance 14.16
	Baseline 0.20	Baseline 0.34	Baseline 0.31	Baseline 0.53	Baseline 2.36	Baseline 2.44	Baseline 0.38	Baseline 0.63	Baseline 2.38	Baseline 3.86	Baseline 7.40	Baseline 0.00	Baseline 1.00	Baseline 6.00	Baseline 1.00
	Signal 0 (100.0%)	Signal 0 (100.0%)	Signal 0 (100.0%)	Signal 1 (100.0%)	Signal 1 (100.0%)	Signal 0 (100.0%)	Signal 1 (100.0%)	Signal 0 (100.0%)	Signal 0 (100.0%)	Signal 1 (100.0%)	Signal 1 (100.0%)	Signal 0 (100.0%)	Signal 0 (100.0%)	Signal 0 (100.0%)	Signal 1 (100.0%)
Evaluation Fields	Outcome 0.43	Outcome 0.50	Outcome 0.69	Outcome 1.88	Outcome 2.93	Outcome 1.67	Outcome 1.63	Outcome 0.38	Outcome 0.75	Outcome 3.14	Outcome 4.00	Outcome 0.25	Outcome 0.33	Outcome 2.00	Outcome 1.00

Appendix 5. Screen shot of the opening page of the HSIP Project Application Tool (Developed by SafeTREC).

* This excel program optimized 1024 x 768 resolution.

Appendix 6: Benefit / Cost Ratio Calculation Method—From Chapter 9: Highway Safety Improvement Guidelines.

Benefit / Cost Ratio Calculation Method

1) $Benefit(Annual) = \dfrac{CRF \times \sum_{S=0}^{4} (N_S \times AC_S)}{Y}$

- CRF : Crash reduction factor in each countermeasure
- S : Severity (0:PDO, 1:Minor Injury, 2:Injuey, 3:Severe Injury, 4:Fatal)
- N_S : Number of selected countermeasure related accident in severity levels
- Y : Crash data time period (Year)
- AC_S : Accident costs in severity levels

Severity	Accident Cost
Fatality (Death)	$4,100,000
Severe Injury	$208,000
Injury – Other Visible	$53,000
Injury – Complaint of Pain	$25,000
Property Damage Only	$2,300

* These crash costs are based on level of severity taken from the National Safety Council's "Estimating the Cost of Unintentional Injuries, 2009" bulletin.

2) $Benefit(Life) = Benefit(Annual) \times (P/A, i, n)$

- $(P/A, i, n)$: Uniform series present worth factor to convert a series of uniform annual amounts to its present value

$$(P/A, i, n) = \frac{(1+i)^n - 1}{i(1+i)^n}$$

i : minimum attractive rate of return expressed as a decimal fraction (i = 0.04)
n : service life of improvement

3) Benefit / Cost Ratio (each countermeasure)

$$Benefit/CostRatio = \frac{Benefit(Life)}{Cost}$$

4) Total Benefit / Cost Ratio

$$TotalBenefit/CostRatio = \frac{\sum_{CM=1}^{3} Benefit(Life)_{CM}}{TotalprojectCost}$$

Appendix 7: Course Template: Methods for identifying corridors where there is a highest potential for reducing pedestrian/bicyclist injury.

Module 1—Introduction

Upon completing this module, participants should be able to have broad understanding of the primary goals and issues involved in developing a network safety plan that maximizes return on investment.

Main Topics

- Concept of maximizing return on investment in the traffic safety planning
- Benefit-cost calculation
- Steps in analysis of locations with high potential for preventing injury

Resources

- Overview of resources listed in modules below
- Overview of traffic safety data/resources in general
- Pedestrian/Bicyclist Information Center site (PBIC)
- SafeTREC pedestrian/bicycle resource site

Module 2—Data Base Construction

Upon completing this module, participants should be able to construct the best possible database for developing a network safety plan using the sources available to them.

Main Topics:

- Infrastructure
- Collisions
- Exposure
- Data structure and quality

Resources (for California)

- Transportation Injury Mapping System (TIMS) http://safetrec.berkeley.edu/tims/index.html (10 years of fatal and severe injury in California, geocoded and downloadable by site, corridor, area, etc.)
- SWITRS
- FARS
- TASAS (for the State Highway System)

Module 3—Calculating "Expected Number of Injuries"

Upon completing this module, participants should be able to identify individual locations, corridors, or areas and calculate their respective expected future injuries.

Main Topics

- Developing set of sites/corridors/areas
- Increasing Years of History and/or Follow-up
- Increasing Scale (from specific sites to corridors, zones, or entire network)
- Combine similar sites
- Bayesian Method

Resources

- Clustering analysis tool (to be developed by SafeTREC)

Module 4—Choosing Potential Countermeasures

Upon completing this module, participants should be able to utilize available resources to identify potential countermeasures for the sites/corridors/areas identified in Module 3.

Main Topics

- Causal factors
- Choosing countermeasures based on causal factors

Resources:

- Pedestrian/Bicycle Information Center (PBIC)
- Walkinginfo.org http://www.walkinginfo.org/pedsafe/
- Bicyclinginfo.org http://www.bicyclinginfo.org/bikesafe/

Module 5—Conducting Benefit-cost Analyses

Upon completing this module course, participants should be able to calculate benefit-cost ratios for the countermeasures identified in Module 4 as mapped on to the sites identified in Module 3.

Main Topics

- Concept of Benefit-cost
- HSIP benefit-cost calculations
- Benefit-cost comparing different countermeasures, different scales of analysis, and different types of collisions (e.g., pedestrian, bicyclist, vehicle only)

Resources

- HSIP Application Tool to Conduct Benefit-Cost Analyses: http://www.dot.ca.gov/hq/LocalPrograms/HSIP/apply_now.htm http://www.dot.ca.gov/hq/LocalPrograms/HSIP/tool_instructions.htm [Note—To use the tool is must be enabled] [Note—Developed by SafeTREC]
- Desktop Reference for Crash Reduction Factors. Federal Highway Administration, Office of Safety. Available at: http://safety.fhwa.dot.gov/tools/crf/resources/fhwasa08011/ (accessed on May 28,
- Crash Modification Factor Clearinghouse. Federal Highway Administration. Available at: http://safety.fhwa.dot.gov/hsip/policy_guide/memo051706.cfm (accessed on May 28, 2011)

Module 6—Strategy for Allocation Across the Network

Upon completing this module, participants should be able to develop an overall network safety plan based on the results produced in the previous steps.

Main Topics

- Mapping resources to treatment options using benefit-cost ratios
- Considerations aside from benefit-cost (e.g., equity, access, mobility)

Resources

- Prioritization tool for strategic resource allocation (to be developed by SafeTREC)